VOLUME 18, 1956/57, No. 7 pp. 181-216 Published 15th December 1956

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

THE LIFE AND RELIABILITY OF VALVES

by K. RODENHUIS, H. SANTING and H. J. M. van TOL. 621.385:621.375.2

In recent years there has been a substantial increase in the use of electronic equipment. Not only have telecommunications developed enormously, but new fields of application have been opened up. Typical examples are industrial measurement and control equipment, electronic computers and many new kinds of navigational aids for ships and aircraft. A modern airliner, with radar, may have as many as 500 valves in its equipment. Moreover, the military use of electronic equipment has expanded on a vast scale. American publications mention figures like 2000 electronic valves in a large bomber and 10 000 on board a battle-ship.

As a result of the rapidly increasing use of electronic valves in industrial equipment, aircraft and etc., far more attention is being paid to the reliability of these components than ever before, particularly in view of the often serious consequences of a valve failure in complex equipment. In the following article new definitions are given for the related concepts "reliability" and "life" A description follows of the measures adopted in the factory, and of those which the user can adopt, to improve the reliability and life of valves.

Orginally, the electronic valve was used predominantly in the field of broadcasting, later extended to include television. While both these applications are still expanding, there has been in recent years an enormous increase in the use of valves in other fields, here referred to quite generally as "professional" applications. These may be classified under four main headings: carrier-telephony, electronic computers, industrial measurement and control equipment, and mobile transmitters and receivers. Several modern "Special Quality" valves for these applications are shown in *fig. 1.*

Owing to the serious consequences that may arise from the failure of a valve in professional equipment, the conventional requirements — for example with regard to the mutual conductance, power dissipation, suitability for wide-band amplification, etc. are no longer adequate. It is evident that valves used for professional equipment should satisfy the additional requirement of reliability, for the equipment should be ready for operation at all times and there must be no risk of sudden failures. Other properties that may be important are long life, shock resistance and narrow tolerances in electrical characteristics, although they may not all be equally necessary in every application. For instance, long life and constant electrical characteristics are required of valves used in computers and in telephone repeater amplifiers, but ruggedness is not so necessary, for in this equipment the valves are not subjected to mechanical vibrations or shocks. On the other hand, the valves used in aircraft transmitters and receivers must be able to withstand severe vibrations and shock, but some variation in characteristics is usually to be tolerated, and life also presents no problems where the equipment is in operation only for short periods a day.

The terms "reliability" and "life"

It is necessary in the first place to define the terms "reliability" and "life". To do this, we postulate an electronic apparatus equipped with a large number S_0 (e.g. 1000) of new valves, all of the same type, and which at a given moment is switched on for a long period of operation. Each valve failure is recorded, so that the number of original valves still functioning (S) can be seen at any instant (t). If S/S_0 is plotted, on a logarithmic scale, as a function of t, a curve is obtained similar to those shown in *fig. 2*, which have all been taken from publications on the subject ¹)²)³)⁴)⁵).

---

¹) N. W. Lewis, Post Office Electr. Engr's J. **41**, Part I, 10-12, 1948.

²) D. K. Gannet, Bell Labs. Record **18**, 378-382, 1940.

³) C. R. Knight, A.I.E.E. Conf. Electron tubes for instrumentation and industrial use, Philadelphia, March 1948.

⁴) Eleanor M. MacElwee, Sylvania Technologist **3**, 16-20, April 1950.

⁵) K. Rodenhuis and W. Sparbier, Elektron. Rundschau **9**, 22-25, Jan. 1955 and 72-74, Febr. 1955.

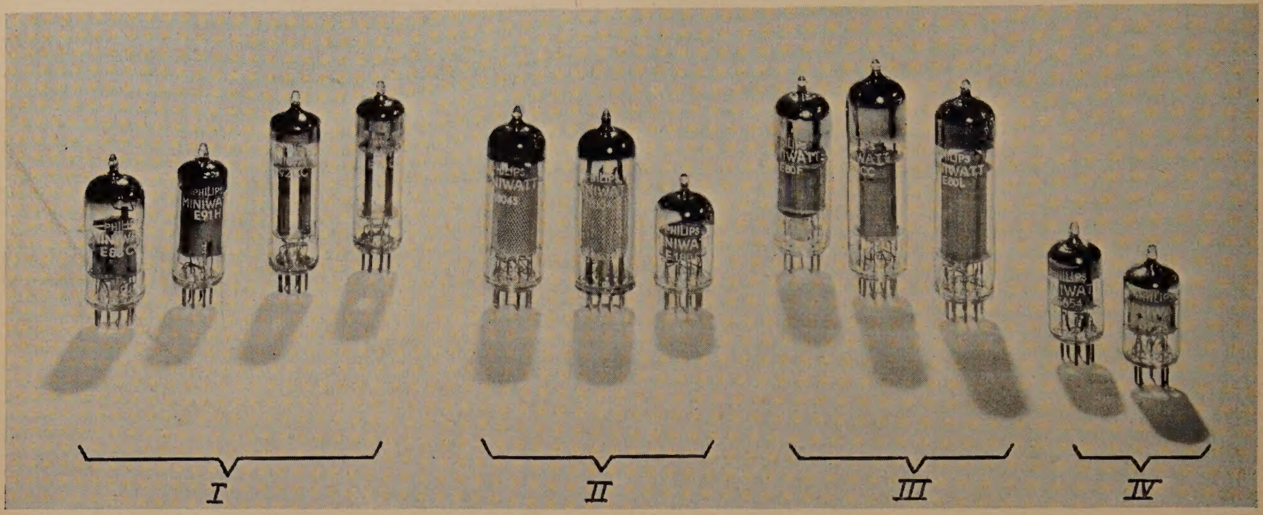

Fig. 1. Some modern "Special Quality" valves for professional equipment.
Group I. Valves for electronic computers (great reliability, long life (also when operated cut-off), narrow spread in characteristics affecting switching functions): E 80 CC, E 91 H, E 92CC, E 90 CC.
Group II. Valves for telephone repeater amplifiers (great reliability, long life, narrow tolerances in amplifying, characteristics): 18045, 18042, E 180 F.
Group III. Valves for control and metering equipment (same characteristics as in group II but also shock-proof): E 80 F, E 80 CC, E 80 L.
Group IV. Valves for mobile radio stations (great reliability, shock-proof): 5654, 5726.

Curve *1* found by Lewis [1]) is noteworthy; it is almost a straight line, having the equation:

$$S/S_0 = \mathrm{e}^{-Pt}, \quad \dots\dots \quad (1)$$

where P is a constant. From this simple expression certain interesting conclusions may be drawn. For example, the number of valve failures per unit time, i.e. $-\mathrm{d}S/\mathrm{d}t$, is $PS_0\,\mathrm{e}^{-Pt} = PS$. Consequently, the percentage failures per unit time (the "failure rate", analogous to the death rate of a population), i.e. $-100\,(\mathrm{d}S/\mathrm{d}t)/S$, is equal to the constant 100 P and is therefore independent of time.

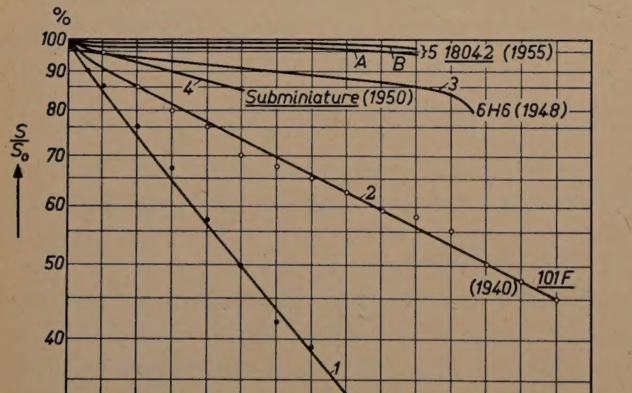

Fig. 2. The percentage S/S_0, on logarithmic scale, as a function of time t, for life tests on various types of valve. The curves *1* . . . *5* have been taken from published sources [1]) . . . [5]); the valve type and the year of publication of the curves are given in each case. The data for curves *5A* and *B* were obtained from a test made with the cooperation of the Netherlands Post Office.

In the case of curve *1*, the average life L_m of the group is found from the equation:

$$L_\mathrm{m} = \frac{1}{S_0}\int_0^\infty S_0\,\mathrm{e}^{-Pt}\,\mathrm{d}t = \frac{1}{P}. \quad \dots \quad (2)$$

For curve *1*, $P = 0.135 \times 10^{-3}$ per hour (failure rate 13.5% per 1000 h) and therefore $L_\mathrm{m} = 1000/0.135 = 7500$ h.

At a time $t = L_\mathrm{m}$, the number of survivors is:

$$S_{L_\mathrm{m}} = S_0\,\mathrm{e}^{-1} = 0.368\,S_0,$$

in other words, 36.8% of the original number are still in operation. At a time $t = 2\,L_\mathrm{m}$, the number of survivors is $S_{2L_\mathrm{m}} = S_0\,\mathrm{e}^{-2}$, and so on: in each period L_m the percentage survivors decreases by a factor e. This means that the valves that have already been in operation for thousands of hours are in no respect to be distinguished from brand-new valves, and that there is therefore no point in replacing them. Obviously, this holds good only for a straight line, such as curve *1* in fig. 2.

Lewis [1]) has pointed out that a straight line is only to be expected where completely random conditions prevail. It cannot therefore be taken as representative of the normal case. The type of curve that should be regarded as normal is, in our opinion, curve *3*, given by Knight [3]). This shows a sharp drop during the first few hundred hours, followed by a long, fairly straight portion with a slight slope, which finally falls steeply again after many thousands of hours. The form of this curve may be explained as follows. In the beginning, some valves show the effects of manufacturing faults not detected in the factory, such as bad welds or near short-circuits between electrodes. After these valves have been removed, there will be merely an occasional "random" failure, until finally certain kinds of predictable defects become prevalent, defects arising from gradual physical and chemical processes in and near the cathode, such as declining emission and diminishing insulation resistance beween electrodes.

A curve of this form (*3*, fig. 2) yields the following information.

1) In the straight central portion of the curve the valves behave more or less in accordance with the exponential law (eq. 1).
2) The higher failure rate during the first 100 to 1000 hours means that valves that have been in operation for several hundred hours are more reliable than new valves.
3) Just past the straight portion, where the failure rate increases, it is advisable (unless certain precautions have been taken, which will be discussed later) to replace all survivors by new valves.

As regards the latter point we shall define the end of a valve's "useful life" (L_p) as *the moment at which the failure rate begins to increase (fig. 3)*. This moment depends, of course, not only on the quality of the valve but also on the circuit, since the permissible deviations that may occur in certain valve characteristics are determined by the particular function of the valve in the circuit. Two important valve-characteristics in this respect are mutual conductance and grid current. To give a typical objective criterion: a valve is regarded as unserviceable when its mutual conductance has dropped to 70% of the nominal value, or when its grid current (under specific conditions) has risen to 1 μA. Other criteria may be chosen for valves used in special equipment, for example in electronic computers.

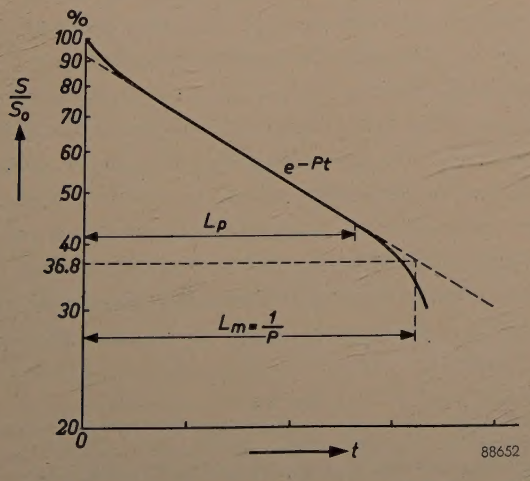

Fig. 3. Illustration of the terms "average life" L_m and "useful life" L_p (L_p comes to an end when the failure rate begins to increase).

Other definitions of "life" have been given by Eleonor MacElwee [4]), who regards it as the time after which 80% of the valves are still functioning, and by Gannet [2]), who takes the average life according to (2) as the life of a valve. Gannet's definition is useful for exponential characteristics (straight lines in fig. 2) but for more representative curves, such as *3*, it gives a result of little practical value. According to MacElwee's definition a "life of 5000 hours" represents in fact a failure rate of about 4% per 1000 h. The present authors prefer to relate the failure rate to reliability rather than to life (see below).

During a practical life test the failure rate (except at the beginning) is constant and relatively low. *This* failure rate mainly determines the frequency of the troubles arising from valve breakdowns. We shall regard the reciprocal value of this failure rate as a measure of *reliability;* the flatter the curve the greater the reliability.

The behaviour of a valve of great reliability is exemplified by curves *5A* and *5B* in fig. 2, both of which apply to a type 18042 pentode, used in telephone repeaters. The two curves refer to different applications. Curve *5A* was plotted from data recorded during a test on a circuit highly sensitive to insulation defects in valves. In this case, each valve was signalled as a failure, via a relay, as soon as its insulation resistance fell to 1 MΩ. Curve *5B* applies to the same type of valve under the same load conditions, but employed in an ordinary AF amplifier where a drop in insulation resistance to 1 MΩ could do no harm. In the right portion, curve *5A* has a slope of 0.5% per 1000 h and curve *5B* a slope of 0.25% per 1000 h. The curves still showed no tendency to bend over even at 10 000 hours.

It is believed that these figures do not represent the utmost that can be reached. It may be expected that with further experience it will be possible to reduce the failure rate of Special Quality valves to about 0.1%, a figure that should represent a reliability sufficient for practically all purposes.

A distinction may be made between the sudden and the gradual failure of a valve. A sudden failure without previous warning may be due to a broken connection, short-circuiting, or a cracked envelope.

A gradual failure implies, for example, slowly declining emission (associated with a drop in mutual conductance and in output power), deteriorating insulation, rising grid current, etc. Gradually failing valves in an equipment can be detected beforehand by regular measurements, and serious breakdowns can be avoided by replacing such valves in good time. The advantage of this "preventive maintenance" as against replacement en bloc is that the valves are replaced only as the need arises. The disadvantage — particularly where the plant concerned is equipped with large numbers of valves — is that it may entail putting the equipment partly or wholly out of operation during the measurements. Methods have been devised, however, which allow the equipment to remain in operation during inspection. For this purpose the working conditions of the valve are deliberately worsened, so that any shortcoming can be more readily detected. In a telephone repeater, for example, the test can take the form of decreasing the heater voltage or increasing the resistance in the grid circuit and then measuring the effect on the anode current; this can be done without taking the valve out of its socket or interrupting its operation. In electronic computers so-called "marginal checking" methods are used, in which it is ascertained by how much the amplitude or repetition frequency of control pulses can be varied before the computer begins to make mistakes.

Efficient preventive maintenance not only forestalls breakdowns but effects considerable economies, since it enables the valves to be used up to the end of their individual lives.

Both types of failures, gradual and sudden, will now be dealt with in turn. A discussion then follows of the contacts between valve and valve holder, electrical tolerances, ruggedness and quality control.

Gradual failures

Cathode phenomena

Most valves have an indirectly heated oxide cathode, which consists of a nickel tube coated with a mixture of crystals of BaO and SrO (and sometimes CaO). A tungsten wire inside the nickel tube heats the cathode to a temperature of 750 to 800 °C [6]). The saturation emission of such cathodes lies between 2.5 and 25 A/cm^2.

A high emission is dependent on the oxide mixture having a low work function, and this is the case only when a small fraction of Ba atoms are present in the oxide mixture (of the order of 0.01%) not combined with oxygen. Free Ba atoms may origninate in two ways:

a) Reaction of the BaO with reducing agents in the nickel. These are minute impurities in the cathode nickel that constitute less than 1% of the total weight but affect the emission properties to a very large extent.

b) Electrolysis of the oxide layer by the cathode current passing through it.

During operation, however, certain processes take place in the valve that tend to decrease the quantity of free barium. If the free barium is removed faster than it is produced the result will be a gradual decline in emission. The processes are of two kinds, physical (evaporation) and chemical ("poisoning").

Poisoning may be due to residual gas. Even with efficient pumping and getter action, a residual pressure of about 10^{-7} mm of mercury remains in the valve. The number of gas molecules in the valve is then 10^{-10} less than at atmospheric pressure, but there are still 3×10^9 per cm^3. Moreover, gas may also be given off gradually by the electrodes and the glass envelope. The gas molecules inevitably come into contact with the cathode — positive ions are in fact attracted by it — and some of these molecules and ions combine with the free barium.

Cathode poisoning may also due to another cause. Various adsorbed compounds — including evaporation products from the cathode — lie at the surface of electrodes, glass and mica. Electron bombardment releases some of these materials: of those reaching the cathode some combine with the free barium.

Another process occurring in the cathode, and which may lead to a valve failure, concerns the nature of the reducing agents in the cathode nickel. In the course of time the reducing agents react with the cathode coating at the boundary between the nickel tube and the coating. One reaction product may be a gradually thickening interface layer of barium orthosilicate (Ba_2SiO_4), which has a high resistance. The hotter the cathode, and the lower the current, the more rapidly the layer grows. This interface resistance effectively adds a resistor in the cathode lead shunted by a capacitance of the order of 10 000 pF; this not only reduces the cathode current but introduces negative feedback, causing a drop in gain, particularly at the lower frequencies [7]).

---

[6]) See for example, R. Loosjes and H. J. Vink, Philips tech. Rev. **11**, 271-272, 1949/50.

[7]) From the extensive literature on this subject we may mention: A. S. Eisenstein, The leaky-condenser oxide cathode interface, J. appl. Phys. **22**, 138-148, 1951, and M. R. Child, The growth and properties of cathode interface layers in receiving valves, Post Office Electr. Engr's J. **44**, 176-178, 1951/52.

The behaviour of these different processes as a function of cathode temperature is very important. At temperatures of 50° to 100 °C below normal, the cathode is much more susceptible to poisoning. The reason is that the supply of free barium to the cathode surface is slowed down at low temperatures, as are diffusion processes in general, and also that certain reactions then take place between barium and other elements that do not occur at normal working temperatures. At higher temperatures, however, evaporation is more pronounced and the interface layer grows faster. There is therefore an optimum cathode temperature, which can be lower the more succesfully cathode poisoning has been combated.

To obtain the best results certain precautions should be taken by the valve manufacturer as well as by the equipment designer (the valve user). We shall first consider the manufacturing side.

a) To reduce *poisoning effects* the configuration of electrodes and screens should be such that no electrons can impinge on glass or mica. The components, moreover, should be kept scrupulously clean, requiring special procedures for cleaning and assembly, such as degreasing, boiling in distilled water, degassing by annealing in vacuum before assembly. Dust-free assembly, long pumping schedules and the use of high-quality getters are also necessary.

b) The *cathode temperature rating* should be as low as possible without increasing too much the susceptibility to poisoning. (There must, of course, be some safety margin to allow for heater voltage fluctuations.)

c) To combat *interface effects*, "passive" nickel should be used for the cathode tube, i.e. nickel containing approximately 0.03% Mg and not more than 0.01% Si (ordinary nickel contains up to about 0.1% Si). It takes longer with passive nickel to activate the cathode, but after 10 000 hours operation at normal cathode temperature there is still no measurable interface resistance.

The equipment designer, or user, can do a great deal to increase the useful life of a valve by observing the following precautions:

a) *Heater voltage.* As a general rule the heater voltage should never be allowed to deviate from the nominal value by more than ±5%. (For this reason, the heaters should not, where it can be avoided, be fed in series.) Where the heater voltage has been stabilized to within 1%, however, the valve can quite permissibly be run at, say, 5% (in some cases even at 10%) below the nominal value.

b) *Bulb temperature.* To minimize the release of gas from the glass the bulb temperature should not be allowed to exceed 170 °C at the hottest point. In this connection, the use of screening cans round the valves calls for some care.

c) *Anode and screen-grid dissipation.* High dissipation causes over-heating of the anode or screen-grid, or both, and also raises the temperature of the cathode, entailing the risk of gas formation and barium evaporation. With regard to poisoning it makes a difference whether a given dissipation is produced by a high voltage and a low current or by a low voltage and a high current. Poisoning effects are more pronounced at high than at low voltages, probably because the compounds adsorbed at the surfaces of anode and screen grid are more readily broken down by high-energy electrons into substances harmful to the cathode. It follows, therefore, that low dissipation at low voltage is advantageous; the permissible cathode current may quite be high.

Insulation faults, grid emission and ion current

Material evaporating from the cathode gives rise to faulty insulation by forming a conducting film on the mica spacers and the glass base of the valve. Deposits formed on the control grid can lead to grid emission. This cause of gradual failure may be remedied by using passive cathode material and by keeping the cathode temperature as low as possible and the control grid cool. The insulation over the surface of the mica spacers can be considerably improved by a coating of magnesium oxide.

When the grid is negative, positive ions in the valve cause some grid current to flow. In a well-manufactured valve with an efficient getter, however, so few positive ions reach the first grid per second that the measured grid current is only 10^{-8} to 10^{-9} A, of which a considerable fraction in any case arises from another cause, namely that the electrodes exposed to electron bombardment emit small quantities of soft X-rays, which release photo-electrons from the control grid. Genuine "ionic currents" (of the order of 1 μA) are found only in valves with an air leak or with an inefficient getter. Both these defects are abnormal and can only be due to undetected manufacturing errors.

The manufacturer can prevent the occurrence of such defects in delivered valves by first keeping them in storage for a month and then checking for the presence of gas current. A very efficient method of checking for air leaks is to apply the *argon test.* If a valve is leaky, a good getter will absorb all constituents of the intruding air except the argon (about 1%). The argon test consists in storing the valves for, say, twenty-four hours in a container filled with argon at atmospheric pressure; after this time the presence of a leak can easily be ascertained by measuring the grid current. In this way, storage for one day is equivalent to a shelf life of 100 days.

The equipment designer can often reduce the effects of imperfect insulation and grid emission by keeping the control-grid resistance as low as possible.

Sudden failures

Sudden failures may be due to broken connections (e.g. a bad weld), short-circuiting between two electrodes, glass cracks and heater-cathode short-circuits.

Broken connections

Practically all open connections in valves occur at or near welds. A reliable welding technique (spot welding) is therefore essential in the manufacture of valves. Non-destructive testing of every individual weld is obviously impracticable, but a generally high standard of welds can be obtained if the following rules are observed:

1) The valve designer should ensure that all welded connections are clearly visible and easily accessible, and that only readily weldable materials are used.
2) The welding equipment should yield a completely reproducible result.
3) Sampling tests on the quality of welds should be carried out at regular intervals.

The second point requires some elucidation. The quality of a spot weld depends on many factors, the most important being the pressure on the weld, the no-load voltage on the transformer secondary, the impedance of the whole secondary circuit, the phase at the moment the current is switched on, the time of the current cycle, the nature of the materials to be welded, the shape and surface of the electrodes, the mass of the moving parts and the atmosphere in which the joint is made. In a modern spot welder (*fig. 4*) all variables are accurately controlled. The optimum pressure for the usual welds is pre-adjusted to a fixed value found experimentally. The secondary voltages for each individual weld are specified in a table; in general, two or three voltages suffice. The operator places the parts to be welded in the appropriate position between the electrodes and starts the weld by depressing a pedal, after which the process is entirely automatic and free of the human element. The transformer primary current flows through two thyratrons in anti-parallel, which open and close the power circuit at the required times. The phase of the initiating current is fixed and entirely independent of the operator. The weld is terminated automatically, usually after one complete cycle of the mains voltage. (One cycle is normally sufficient, but in some cases several cycles may be required, which are here obtained by means of a control circuit using cold cathode valves.) The welding electrodes, made of a copper alloy, are exchanged at regular intervals, the old ones being reshaped on a special milling machine. A stream of a mixture of hydrogen and nitrogen is played on the weld to prevent oxidation. The mass of the moving arm is small, so that its inertia causes no significant variation of the electrode pressure as the materials compress during the welding operation. All these measures ensure reproducible welds of a consistently high quality.

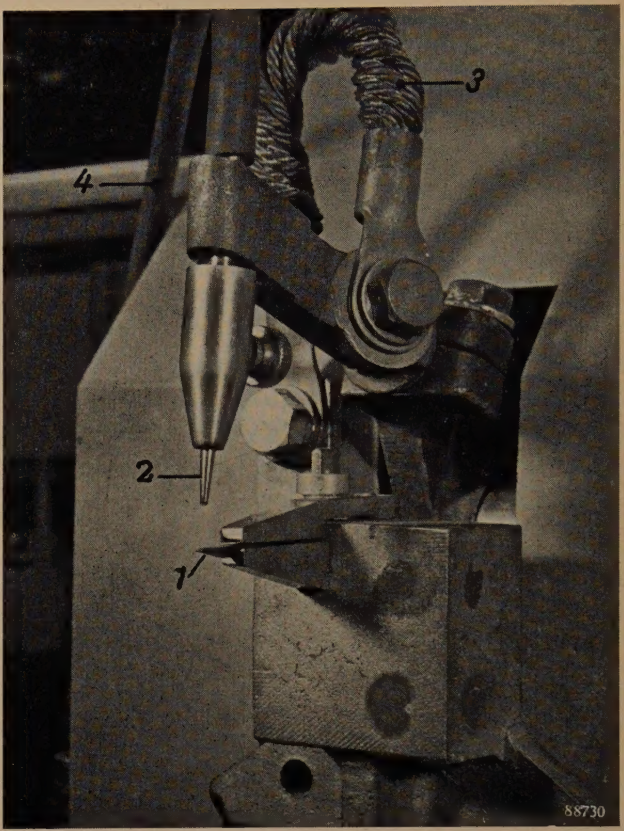

Fig. 4. Spot-welder. *1* fixed electrode, *2* moving electrode, *3* current supply cable, *4* hose through which a stream of a reducing mixture of H_2 and N_2 is played on the weld.

Short-circuits

Short circuits due to direct contact between two electrodes (which may be only 100 μ or even 50 μ apart) are easy to detect during production and are therefore extremely rare in the finished product. A much more difficult problem is the prevention of partial short-circuits or conducting paths arising from foreign particles in the valve. "Short-circuits" of this nature may show values of resistance varying from 10^7 to 1 ohm. Moreover, they may only be intermittent, which makes them very difficult to detect.

These particles originate in many ways. In the first place they may be present in the air as dust, and therefore the air supply to the assembly shop is always filtered. The formation of dust in the assembly shop itself is reduced by using materials that give off little dust for walls, ceiling, benches, curtains, etc. Floors and benches should be repeatedly wiped down with wet cloths. Well-polished linoleum is a good dust-trap. The parts are assembled and welded under dust covers, screened off at the front by nylon curtains as shown in *fig. 5a* and *b*.

The components themselves, of course, must also be free of dust. They are therefore produced under conditions that are as dust-free as possible, after which they are subjected to washing and stoving processes before being conveyed, suitably packed, to the assembly shop. Only glass, metal or plastics can be used as packing material. For the conveyance and storage of grids, for example, a metal packing has been developed in which the grids can remain during the washing and stoving processes.

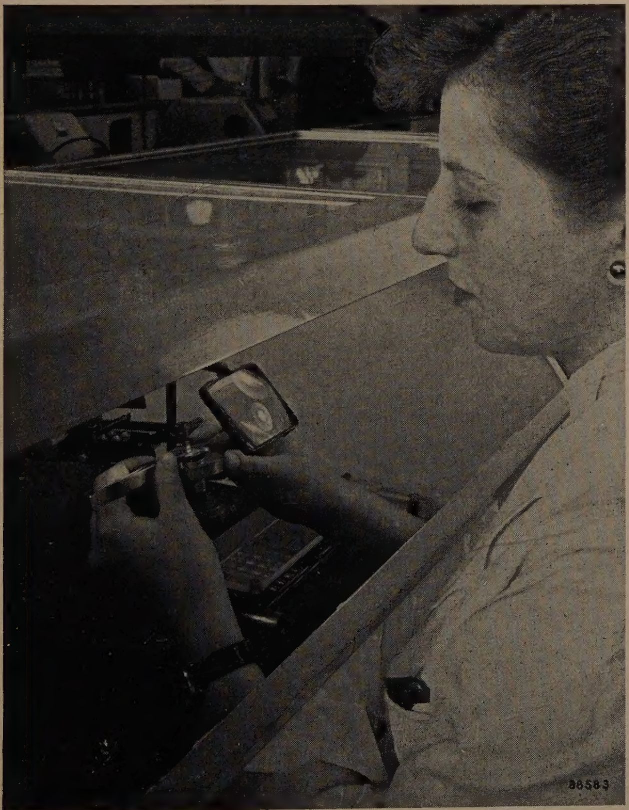

Fig. 5*a*. Assembling the electrode system of an amplifier valve. This operation, like that in fig. 5*b*, is carried out under a transparent dust-cover. The operator's hands enter the dust-cover through slits in a nylon curtain, which produces no dust itself.

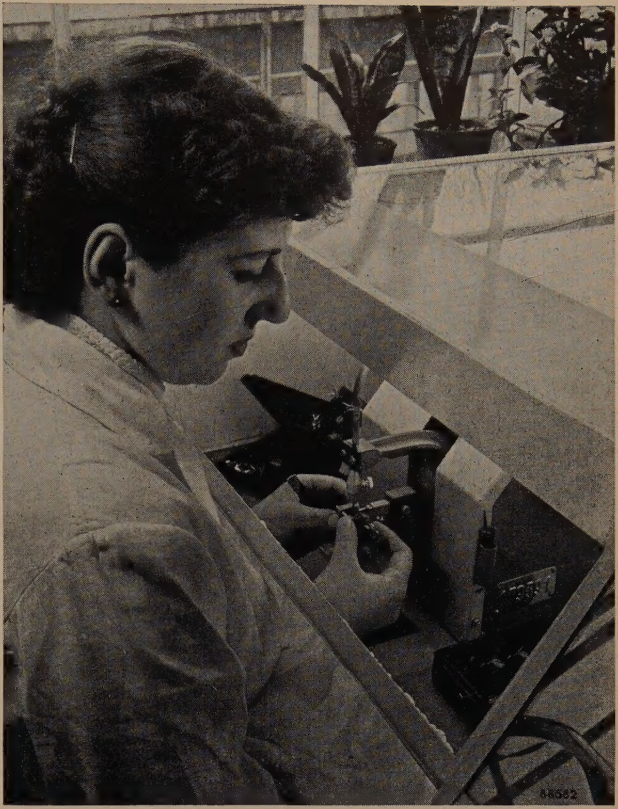

Fig. 5*b*. Spot-welding the electrode system of an amplifier valve.

Finally there are the particles introduced during production: metal particles may fly off during welding, the coating of the cathode may be chipped and bits may flake off the mica spacers (mica particles become conductive in the long run under a coating of cathode material). Sputtered metal particles point to faulty adjustment of the welding machine, and chipped cathode material indicates a lack of care in the manufacture of the cathode. Formation of mica particles can be reduced by using a new form of spacer, which flakes much less than the old type (*fig. 6a* and *b*).

In spite of all these measures, the quality of a valve is still to some extent dependent upon the girl at the assembly bench. It is not enough to provide her with the best tools and to simplify and reduce the fatigue of the operations she has to perform. It is equally necessary to ensure that her working environment is fresh, clean and orderly as befits the quality expected of the product. What is more, the operatives should understand the importance of their work; this should be explained to them during their training period. Last but not least, the results are considerably influenced by the wage policy. In our opinion the best method is to pay a piece-rate in combination with a system of bonuses based on the quality of the work performed.

The quality is expressed in so many marks out of ten, the marks being given, in accordance with a fixed schedule, after making a number of sample tests.

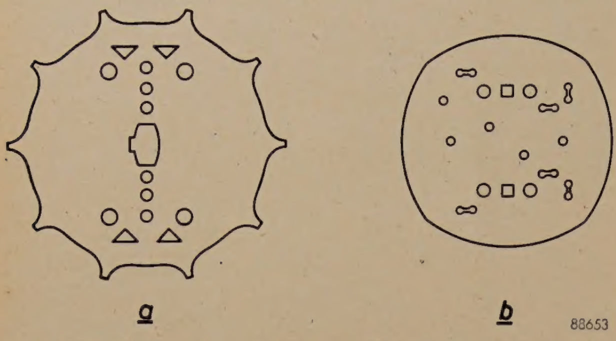

Fig. 6. Old form (*a*) and new form (*b*) of mica spacers. The new form reduces the incidence of flaking and chipping.

Glass defects

After the electrodes have been mounted on metal rods, which are fused into a glass base, the glass base is sealed into the envelope. This is done in an automatic sealing machine, which must be checked constantly on its performance in order to prevent the occurrence of glass strains which might ultimately lead to cracks and hence to sudden failures. The variables here are the gas jets, the intensity and height of which depend upon the composition and pressure of the gas, and the glass, which is not a perfectly constant product. Each production run is preceded by the sealing in of a number of dummies, which are inspected for glass strain with the aid of a polariscope [8]), and for the form of the seal by making a cross-section. After the machine has been adjusted for optimum sealing, the production run is started. Every hour samples are taken: after cooling, the sample valves are immersed in boiling water and a metal cone of prescribed form is pressed between the pins; they are then plunged directly into cold water. The occurrence of cracks gives indications of how the machine should be readjusted. After exhausting, all valves are immersed for a moment in boiling water. This test should result in no rejects; a reject is a sign that the sealing machine needs readjustment.

Heater-cathode short-circuits

One type of failure that is due to a gradual physico-chemical process and yet occurs suddenly, without any warning, is the breakdown of the insulation between heater and cathode [9]). The nickel tube of an indirectly heated cathode is heated to a temperature of 750-800 °C by a tungsten (usually helical) filament, which itself has a temperature of about 1100 °C (*fig. 7a*). The only material that has been found suitable at this temperature to insulate the heater from the cathode, while preserving adequate adhesion and flexibility, is alumina (Al_2O_3). The tungsten is coated with alumina by spraying or by electrophoresis [10]), and afterwards sintered at a high temperature.

In most cases the heater-cathode insulation does not have to satisfy very stringent requirements. For example, in a high-frequency amplifier, with the heaters connected in parallel, the potential difference between the cathodes and the middle of the heater is only a few volts, viz. the voltage drop over the cathode resistor. It is quite another matter, however, in circuits where the heaters are connected in series or where the cathode is used as an output electrode and is at a high potential with respect to earth. There is then a danger that the insulation will ultimately break down owing to gradual electrolysis of the alumina at those points where it makes direct contact with the cathode nickel. When the heater is positive with respect to the cathode (the other polarity will be considered presently) negative oxygen ions are drawn to the heater where they cause oxidation. The tungsten oxide thus formed dissolves in the alumina, producing an aluminium tungstate (fig. 7*b*). The resistance of this compound is much lower than that of alumina. After a time determined by the voltage V_{kf} and the temperature of the heater the insulation between

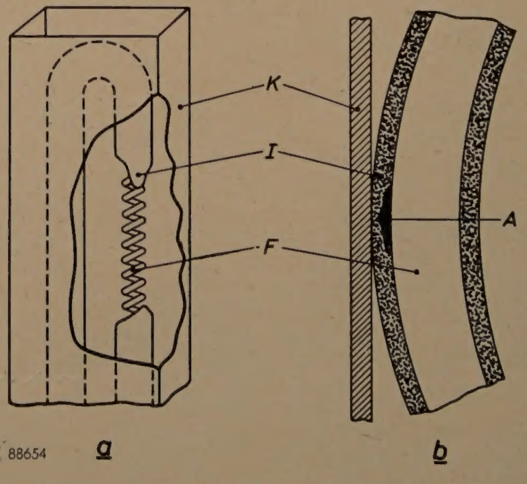

Fig. 7. *a*) Indirectly heated oxide cathode. *K* nickel cathode tube, *F* tungsten heater, surrounded by an insulating layer of alumina *I*.
b) If the heater is positive with respect to the cathode, a conducting layer of an aluminium tungstate (*A*) forms in the alumina coating at the points where the latter touches the cathode nickel, which may lead to a heater-cathode short-circuit. (Drawings not to scale.)

---

[8]) Philips tech. Rev. **9**, 277-284, 1947/48; **14**, 290, 1952/53.

[9]) This phenomenon has been investigated by P. G. van Zanten and P. N. Kuiper of the Physico-chemical Laboratory, Electronic valve factory, Eindhoven.

[10]) S. A. Troelstra, Philips tech. Rev. **12**, 293-303, 1950/51.

the cathode and a point of the heater will suddenly break down, possibly resulting in a complete short-circuit. If the impedance in the heater-cathode circuit is low, the short may lead to the fusing of the heater.

Fig. 8 shows the relationship between V_{kf} (heater positive) and the time t in which 1% of a batch of valves develop heater-cathode shorts, with the heater voltage as parameter. (The values are for guidance only, having been obtained by extrapolation of accelerated tests carried out at higher voltages and temperatures.) The figure shows clearly how the life of a valve is shortened by raising V_{kf} and how this is influenced by the heater voltage. It can be seen that V_{kf} should not exceed 100 V if a life of 10 000 hours is required.

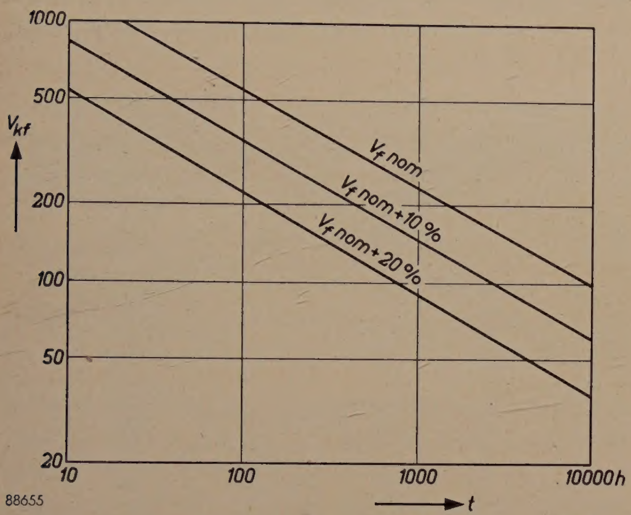

Fig. 8. Potential V_{kf} (heater positive with respect to cathode) as a function of the time t in which 1% of the valves show heater-cathode short-circuits, at nominal heater voltage $V_{f\,nom}$ (heater temperature 1080 °C) and at 10% and 20% higher heater voltages. Provisional curves only, obtained by extrapolation of accelerated tests.)

The situation is less serious if the cathode is positive with respect to the heater. In this case it is the nickel that is oxidized, but since nickel is not so readily oxidized as tungsten, the process takes much longer (about ten times longer). This means, however, that to obtain the same life as with the other polarity the value of V_{kf} may only be about twice that of the former case.

To counteract this phenomenon the manufacturer can try to prevent contact taking place between the cathode and the heater insulation touching the inside of the cathode, or at least to allow contact only via a long insulation path. This is done for instance in booster diodes for television receivers, in which peak voltages of many thousands of volts occur between cathode and heater. However, the method is only practicable for large cathodes, and has the drawback of requiring a higher heater temperature, which tends to make heaters brittle. In many Philips valves for professional use the method adopted for smaller cathodes is to coat not only the tungsten but also the inside of the cathode tube with a layer of alumina. This has two advantages:

a) The temperature of the inside coating is lower than the coating on the tungsten, which offers a greater safeguard against breakdowns.
b) The heat transfer is improved by the higher radiation absorption factor of alumina compared with nickel, which means that the heater temperature can be lowered.

It has been found that valves with cathodes treated in this way (types E80CC and E80L are examples) function 5 to 10 times longer before insulation breakdowns occur between heater and cathode, or that, for the same life, V_{kf} can be appreciably higher.

Insulation breakdown due to electrolysis currents as described above is in no way connected with the leakage currents, often much larger, that may flow between heater and cathode. Such currents may be due to emission phenomena (e.g. the presence of BaO on the tungsten or on the inside of the cathode) or to ionic conduction arising from impurities (e.g. sodium) in the alumina coating.

The contacts between valve and valve holder

The quality of base pins and socket contacts, by which a valve is connected with the circuit, has a very important bearing on the quality and reliability of the whole. An investigation conducted by Morrell[11]) of the British General Post Office demonstrated that changes in level (even interruptions) and noise in long-distance telephone communications are due in large measure to bad contacts, including base-pin contacts. Morrell concludes that the only answer is to solder the valves into the circuit.

In our opinion this drastic step is acceptable only where the life of the other circuit elements is of the same order as the valve life, or shorter. Usually, however, equipment life is many times longer than the average life of the valves, so that the replacement of soldered-in valves would create a serious maintenance problem. On the whole, new valves and new valve-holders make excellent contact; the contact deteriorates in the course of time owing to corrosion. Only precious metals maintain stable contact under corrosive conditions, a fact which has been proved by the excellent results obtained from a series of extensive tests carried out, under very adverse conditions, with gold-plated pins and socket contacts. The majority of Philips "Special Quality" valves are now equipped with gold-plated base pins.

---

[11]) F. O. Morell, paper read before the London branch of the Institution of Post Office Electrical Engineers, 5 April 1948.

Spread in characteristics

Discrepancies in electrical characteristics between valves of the same type are attributable to differences in the dimensions or clearances of electrodes, in the work function of the control grid and in cathode emissivity.

Pentode type 18042 will serve to illustrate the effect of dimensional variations on the electrical characteristics. This valve has clearances of 120 μ between cathode and first grid and 310 μ between the first and second grids. The following table shows the change in grid voltage V_{g1} and mutual conductance S at 10 mA anode current (screen-grid potential 120 V) for 1μ variations of various electrode dimensions.

	Change in V_{g1} mV/μ	Change in S (mA/V)/μ
Cathode diameter	10	0.06
Grid 1 diameter	1.5	0.06
Grid 1 pitch	90	—
Grid 1 wire diam.	115	—
Grid 2 diameter	5	—

It is evident that special measures are needed in the manufacture of this type of valve in order to satisfy the specification, say, that V_{g1} should not vary by more than ±330 mV from the nominal value (as required for the pentode 18042). It calls, among other things, for extremely narrow tolerances in the diameter and tensile strength of the wire for the first grid, and for the utmost accuracy in the cathode diameter (the grid pitch can be set precisely relatively easily).

The work function of the first grid in a new valve may fluctuate initially, but will settle down if the valve is allowed to burn in for 48 hours. The work function depends largely upon the extent to which evaporation products from the cathode have become deposited upon the grid surface, that is to say it depends on the cathode temperature. The grid is more susceptible to poisoning than the cathode since it does not itself produce free barium but can only receive it from the cathode. The aim of the 48 hour ageing period is therefore to allow the surface of the grid to reach its normal working condition, after which there will be no appreciable change in the work function.

In discussing electrical tolerances, mention should be made of a new grid construction, evolved elsewhere [12]). Until quite recently, grids were almost invariably made as shown in *fig. 9a*. Molybdenum grid wire is wound helically around two "backbones" made of nickel, copper-clad wire or a similar material, in which grooves are cut to hold the wire turns in place, the grooves being closed under pressure after winding is completed. The grid wire itself, whose turns may be oval or round in form, have not only an electrical function but also a mechanical function, in that, together with two mica spacers, they determine the spacing between the backbones. In the new construction [12]), illustrated in fig. 9*b*, the grid wires have a purely electrical function. The mechanical function is fulfilled by a sturdy frame consisting of two molybdenum rods connected, with the right separation by four spot-welded molybdenum strips. Very thin tungsten wire (e.g. 7.5 μ) is wound tightly around the frame, the winding tension corresponding to about 60% of the tensile strength of the wire. The wire is fixed to the frame with gold solder or with molten glass powder. The thickness of these frame grids, that is to say the clearance between the turns on opposite sides of the frame, is wholly determined by the thickness of the molybdenum rods, which can be manufactured within a precision of 5 μ.

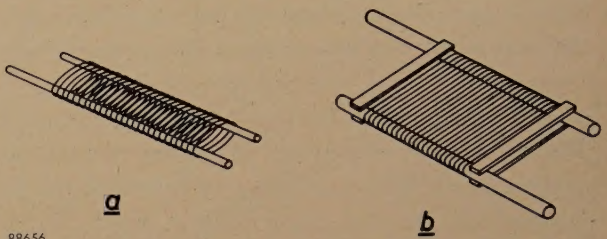

Fig. 9. A grid of normal construction (*a*) and a modern "frame grid" (*b*).

In conjunction with measures to reduce the tolerance in the dimensions of the cathode and the mica spacers, the use of a frame grid allows valves to be made with a cathode-grid clearance of 50 μ, that is, half the clearance that can be reached satisfactorily with normal grids. Several types of Philips valves have already been fitted with frame grids (e.g. E180F and E88CC, shown in fig. 1) and other types will follow in the near future.

Shock and vibration resistance

The enormous increase in the use of electronic equipment in aircraft has made it essential to make valves that can withstand severe vibrations and shock.

---

[12]) G. T. Ford, Bell Labs, Record **27**, 59-61, 1949. E. J. Walsh, ibid. **28**, 165-167, 1950. G. T. Ford and E. J. Walsh, Bell Syst. tech. J. **30**, 1103-1128, 1951.

If there is the slightest play between the mica spacer and the glass envelope, or between the mica and a component which it holds in place, the mica will begin to wear as soon as the valve is exposed to vibrations and this will tend to increase the amount of play. The first of several harmful consequences will be increased microphonic noise. In an advanced stage the play may cause shorts between two or more electrodes. Moreover, the mica may release gas, and mica flakes may come in contact with the cathode and poison its surface. Mechanical fatigue may lead to rupture, particularly of welds, filament and cathode connecting strip.

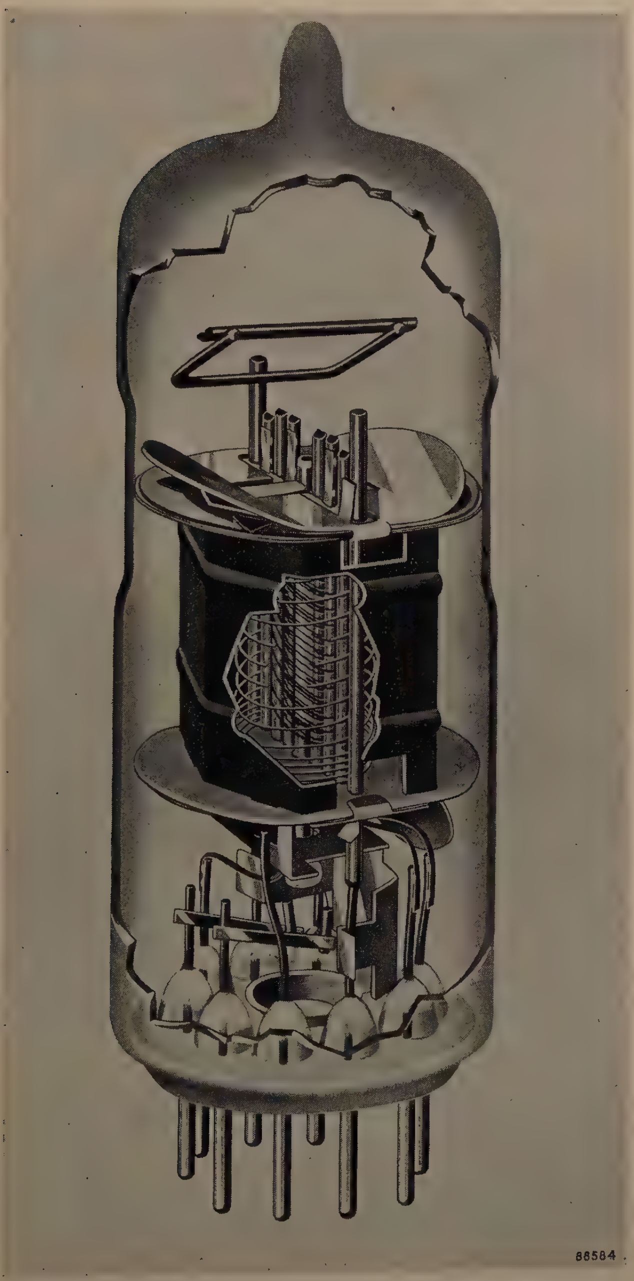

Fig. 10. Type E80F pentode, fitted with mica spacers as in fig. 6*b*. The upper spacer fits tightly in a constricted region of the envelope.

To improve a valve's ability to withstand adverse mechanical conditions it is first of all necessary to restrict the forces acting on the components. This is done by using the lightest possible components and by avoiding mechanical resonances under 200 c/s, and preferably under 1000 c/s. The toothed mica spacers formerly in general use (fig. 6*a*) are liable to break off at the teeth under vibration (especially in large valves) which gives rise to play and to flaking mica. The new form of spacer illustrated in fig. 6*b* is a substantial improvement. An especially good fit is obtained by very accurately narrowing the bulb at the height of the spacer to an inner diameter about 0.1 to 0.2 mm smaller than the largest diameter of the spacer. *Fig. 10* shows the interior of a valve in which this method has been employed. Mechanical fatigue of the heater can be prevented by restricting the movement of this element inside the cathode tube. The cathode connector can be safeguarded against rupture by giving it a large bend.

Various methods have been devised for testing valves under adverse mechanical conditions. The American military specifications, which are used as standard for Philips valves, prescribe a vibration test lasting in total 96 hours, divided into three equal periods of 32 hours, during each of which the acceleration force is applied in one of the three main axes of the valve. The specified repetition frequency is 25 c/s and the peak acceleration 2.5 *g* (*g* being the acceleration due to gravity), and the amplitude 1 mm. The American specifications also lay down a shock test consisting of five shocks of 500 *g*, each lasting 1 millisecond, and applied in the direction of each of the axes; for the longitudinal axis five blows are given in the one direction and five in the opposite direction. This test is carried out on a machine of American design. (*fig. 11*).

The valves are considered to have passed these tests when they are still in working order and when their electrical characteristics still fulfil certain requirements. As the tests are, in a sense, of a destructive nature, they can only be carried out by sampling. The American military specifications therefore lay down sampling plans [13]) which are based on an "acceptable quality level" of 6.5% rejects for the vibration test and 20% for the shock test. The figures indicate that valves which have satisfied these specifications should not be exposed in practice to vibrations and shocks of the same magnitude as during the tests, as otherwise the failure rate would be very high.

---

[13]) For a general treatment of sample testing and sampling inspection plans, see Philips tech. Rev. **11**, 176-182, 260-270 and 362-370, 1949/50.

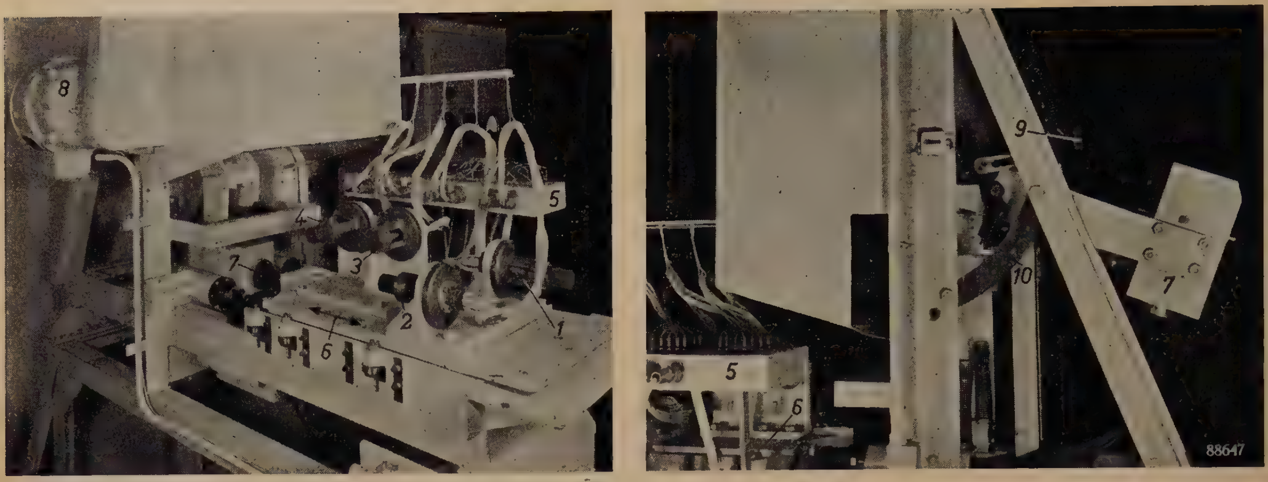

Fig. 11. Two views of a machine for shock-testing valves in accordance with American military specifications. *1*, *2*, *3*, and *4* are valve bases mounted facing different directions, connected to the supply rack *5*, and fixed rigidly to a carriage *6* which can slide in the directions indicated by arrows. A hammer *7*, driven by a motor *8*, strikes the carriage after a fall from a height adjusted by screw *9* along a scale *10*. After 5 consecutive blows the positions of the valves are changed until they have passed through all four positions *1* . . . *4*. They are then removed and inspected for damage.

Quality control

In the manufacture of "Special Quality" valves, the methods of quality control play a very important part. Some examples have already been mentioned in connection with assembly work and glass strains. However, these methods start as far back as in the manufacture of components, where samples are tested on their most important properties. After pumping and "screening" (cathode activation and stabilization), the valves are tested for short-circuits, open connections, crackling, etc., and certain of their electrical characteristics are measured. They are then subjected for five minutes to a vibration test in order to detect the presence of loose particles, and this is followed by a 48 hours' ageing period to stabilize the electrical characteristics, which constitutes a short life test on every valve produced. Rejects occurring during this period give valuable indications concerning the reliability of the valves. After ageing, all valves are tested on their more important characteristics, such as emission, mutual conductance and anode current and examined again for crackling and insulation. Each week's production is stored separately. From each batch several samples are taken, some being subjected to measurements on capacitances, grid emission, noise, microphony, etc. and others to a life test of 500 hours.

The results of these sample tests determine whether the week's production is suitable for delivery. Before dispatch, the valves are tested once again on their principal characteristics in order to ascertain whether they have suffered during storage.

Conclusion

The purpose of this article has been to give some idea of the problems confronting the manufacturer of special quality valves for professional equipment, and also to indicate how the valves can be employed to best advantage in electronic equipment. As stated, a failure rate of 0.5% per thousand hours with a useful life of 10 000 hours does not mean that endeavours to improve the reliability of these valves are at an end. Much work undoubtedly remains to be done before the failure rate can be reduced to 0.1% per 1000 hours. The equipment designer can help in reaching this goal by making available the extensive practical data at his disposal; close cooperation in this respect between valve manufacturer and valve user can do a great deal to improve the quality of this class of valves.

Summary. The greatly increased use of valves for professional equipment has focussed the attention of manufacturers and users alike on the reliability and life of these components. The authors define "reliability" as the reciprocal of the failure rate, i.e. the percentage failures in a batch of valves after 1000 hours operation. They regard the end of a valve's useful life as the moment at which the failure rate begins to increase. A useful life of 10000 hours and a failure rate of 0.5% per 1000 hours have already been reached with "Special Quality" type valves; it is hoped to be able to reduce the failure rate to 0.1% per 1000 hours.

Valve failures may be either gradual or sudden. The causes of both categories of failure are discussed separately and a description is given of the measures taken in recent years to reduce the incidence of valve defects. The spread of characteristics and shock and vibration resistance are also discussed. In conclusion the desirability of close cooperation between valve manufacturer and valve user is underlined.

A FLYING-SPOT SCANNER FOR TELEVISING 35 mm FILM

by F. H. J. van der POEL.

621.397.611.2:778.53

Flying-spot scanners are particularly suited for the televising of flat, transparent pictures. In a previous article in this Review a flying-spot scanner for televising stationary objects was described. A similar scanner has now been developed for the transmission of 35 mm cinefilm, and the present article discusses the problems involved in scanning a succession of images.

Introduction

The transmission of films has become a feature of every television programme. The equipment used for this purpose is required to derive an electrical signal, suitable for television, from standard film made for cinematographic projection. Such films are recorded at 24 frames per sec; the European television standards, however, require 25 pictures per sec. Since there is no simple method of obtaining 25 television pictures per sec. from 24 film frames per sec, the obvious expedient is to match the film speed to the speed required for television. Increasing the film speed from 24 to 25 frames per sec has no objectionable effect either on the picture or on the sound. In this respect Europe is better off than the United States and other countries where the mains frequency is 60 c/s, and where the number of television pictures per second is accordingly fixed at 30; the film speed cannot be raised so high without making movements appear too fast and music and speech too high-pitched. In their case, therefore, the solution is more complicated.

Every television picture is formed from a raster consisting of a certain number of horizontally scanned lines. According to the Gerber standard [1]) which is used for example, in the Netherlands, the number is 625. Moreover, the scanning lines are interlaced, which means that the odd lines 1, 3, 5, etc. are traced first, and then the even lines 2, 4, 6, ... The vertical deflection of the light-spot which traces a television raster is shown in *fig. 1a* as a function of time. Each complete picture period T of 40 milliseconds ($^1/_{25}$ sec) may be broken down into four intervals: the raster flyback times T_{b1} and T_{b2} are each 1.5 msec, leaving 18.5 msec each for the actual scanning periods T_{s1} and T_{s2}. Strictly speaking, this means that the number of lines per odd or even raster is not exactly 625/2 but only (18.5/40) × 625.

For the televising of films, frequent use is made of flying-spot scanners. The interlacing, particularly, gives rise to a number of problems. In the following description of a flying-spot scanner for 35 mm film developed in Philips Research Laboratories, the solution of these problems will be discussed. Before this we shall first touch briefly on a system of televising films with the aid of a normal television camera.

Televising of films using a television camera tube

A familiar method of televising films is that in which a normal storage-type camera tube (iconoscope, image iconoscope) is used in conjunction with a slightly modified conventional cinema projector. The film is fed intermittently through the projector in the normal way, but at a rate of 25 frames per second. Each film frame is projected on to the light-sensitive layer in the camera tube during the brief flyback periods T_{b1} and T_{b2} (fig. 1*b*). The potential

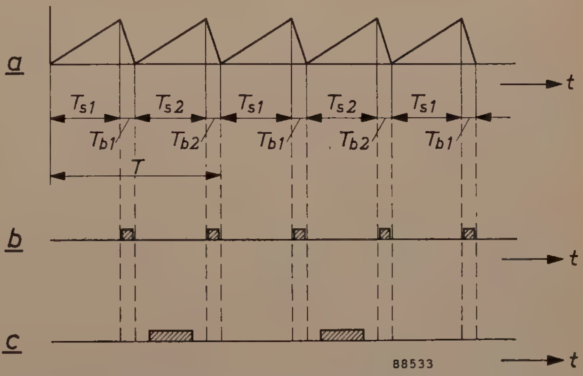

Fig. 1. *a*) Vertical deflection, as a function of time *t*, of the spot tracing a television raster. T = picture period (40 msec). The "odd" and "even" lines of the raster are scanned during the intervals T_{s1} and T_{s2} respectively (each 18.5 msec). During the flyback intervals T_{b1} and T_{b2} (each 1.5 msec), when the spot returns from the end point of one raster to the starting point of the next, the spot is suppressed.
b) Times during which the film frames are projected when televising a film with a normal television camera tube. These times fall within the intervals T_{b1} and T_{b2}.
c) The film is moved forward one frame during the T_{s2} intervals.

---

[1]) See, e.g. W. Werner, The different television standards considered from the point of view of receiver design, Philips tech. Rev. **16**, 195-200, 1954/55.

pattern thus produced on the target of the camera tube is then scanned during the intervals T_{s1} and T_{s2} by an electron beam[2]. During scanning the projected light beam is intercepted by a shutter. In this way use is made of the camera tube's "memory", in which the potential pattern remains for a while after the exposure has ended.

The advantage of this system, in that it allows a normal studio television camera to be combined with a conventional cinema projector, is offset by the fact that storage-type camera tubes give rise to signals even when not illuminated[3]. During illumination these spurious signals are superimposed on the desired signal. Spurious signals are not troublesome if there is sufficient intensity of illumination, but this is by no means always the case in the televising of films.

With the development of camera tubes of the vidicon type, in which the drawback of spurious signals has been largely overcome, the system employing camera tubes has recently been gaining ground[4], particularly since the equipment needed is cheaper. At the present time, however, in Europe at least, flying-spot scanners are in general use and generate a television signal of very good quality.

Principle of the flying-spot scanner

As described in a previous article[5], a flying-spot scanner generates a well-defined and reproducible television signal from a flat, transparent object. The principle is briefly as follows.

A point source of light of constant intensity, formed on the screen of a specially designed cathode ray tube (the scanning tube) is projected by an optical system on to the transparent object — the film. (In the following we shall denote the point of light on the scanning tube as the *source* and its projection on the film as the *spot.*) The source is made to trace on the scanning tube a raster of the required number of lines, so that its image, the spot, traces an identical raster on the film. The light passing through the latter is concentrated on a photoelectric cell by means of a condenser lens, so that a photo-current is generated which at every moment is proportional to the transparency of the film at the point defined by the spot. The current passes through a resistor, the voltage across which — after amplification — represents the required picture signal.

The film movement

A fundamental difference between the method of televising films with a flying-spot scanner and the method employing a television camera tube is that in the former it is the film image itself which is scanned, whereas in the latter it is the potential pattern formed in the camera tube. This means that, with a flying-spot scanner, the film image must be available for scanning during both intervals T_{s1} and T_{s2}: it is not possible to use the interval T_{s2} for moving the film one frame farther, as is done in the camera tube method (fig. 1*c*). A process might be imagined in which the film is transported during the flyback interval T_{b2} and kept stationary for the rest of the picture period. In that case, however, a film frame would have to be moved forward in 1.5 millisec, which would entail considerable mechanical difficulties. (In cinema projection it takes about 10 milliseconds to move a film one frame forward.) It has not yet proved possible to construct a 35 mm projector capable of such rapid film transport; even for 16 mm film, where the accelerations are so much smaller, the problem has not yet been satisfactorily solved.

However, it is not a prerequisite that the film should remain stationary during scanning. What is essential is that there should be a certain definite relative movement between the film and the spot, which can equally well be effected partly by the film and partly by the spot. In the flying-spot scanner to be discussed the film is transported at a *constant* speed, corresponding to 25 frames per second. We shall assume that the film velocity v is directed vertically downwards. The spot is made to move such that during each scanning period it has a velocity component w directed vertically upwards. Film and spot then have a relative vertical velocity of $v + w$. In a scanning period T_s ($= T_{s1} = T_{s2}$) the spot has therefore moved over a distance $(v + w)\ T_s$ with respect to the film. The value of w is chosen such that

$$(v + w)\ T_s = h \quad . \ . \ . \ . \ . \ . \ . \quad (1)$$

in which h is the height of the part of the film frame to be scanned. The film is threaded in such a way that at the beginning of a scanning period T_{s1} (fig. 1*a*) the spot starts at the bottom of a film frame, and then scans the odd lines during this period (fig. 2*a* and *b*). After the flyback period T_{b1} the spot begins to scan the even lines, which should fall exactly between the lines scanned during T_{s1}. However, as may be seen from fig. 2*c*, this is not the case; line *2*, for example, does not fall between lines *1* and *3* but half a frame higher. This spot is therefore intercepted by a shutter during the scanning of the even raster. How the even raster is actually produced will be described in the next section. The scanning of the odd lines, however, presents no further problems; when the spot begins, $^1/_{25}$ sec later, to trace the odd lines again, the film has moved exactly one frame farther (fig. 2*d*).

---

[2] See P. Schagen, H. Bruining and J. C. Francken, The image iconoscope, a camera tube for television, Philips tech. Rev. **13**, 119-133, 1951/52. and J. C. Francken and H. Bruining, New developments in the image iconoscope, Philips tech. Rev. **14**, 327-335, 1952/53.

[3] See page 123 of the article by Schagen, Bruining and Francken referred to under [2].

[4] H. N. Kozanowski, Vidicon film-reproduction cameras, J.S.M.P.T.E. **62**, 153-162, 1954; R. G. Neuhauser, Vidicon for film pick-up, J.S.M.P.T.E. **62**, 142-152, 1954; W. L. Hurford and J. Marian, Monochrome vidicon film camera, R.C.A. Review **15**, 372-388, 1954; also L. Heyne, P. Schagen and H. Bruining, An experimental photoconductive camera tube for television, Philips tech. Rev. **16**, 23-25, 1954/55.

[5] A flying-spot scanner developed in Eindhoven is described by F. J. H. van der Poel and J. J. P. Valeton, The flying-spot scanner, Philips tech. Rev. **15**, 221-232, 1953/54 and A. Bril, J. de Gier and H. A. Klasens, A cathode ray tube for flying-spot scanning, Philips tech. Rev. **15**, 233-237, 1953/54.

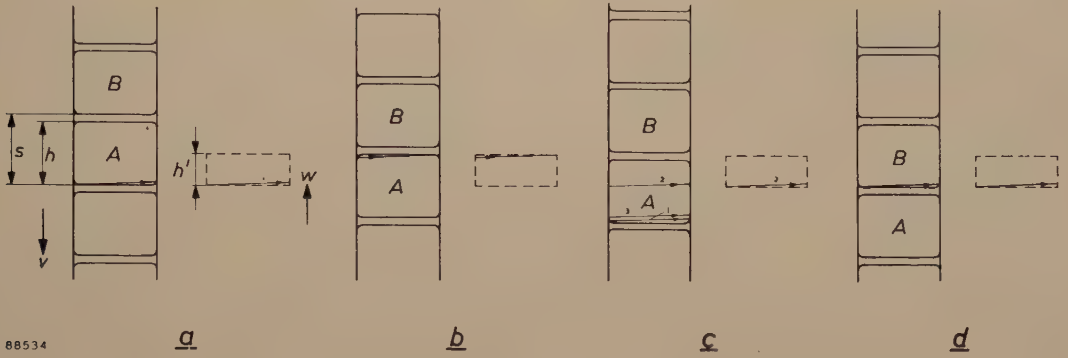

Fig. 2. *a* and *b*) The situations at the beginning and end of an "odd" raster scanned on film frame *A*. For clarity the outline of the raster is shown at the side of the film. v = film velocity; w = vertical velocity of scanning spot; s = distance between two frames; h = height of part of film frame to be scanned; h' = height of raster in film plane.
c) For the beginning of the "even raster, the scanning spot starts at the wrong position on the film (line 2 should fall between 1 and 3); it therefore has to be cut off.
d) After a full picture period the situation is again as in (*a*), but the film has moved forward one frame.

Since the film movement provides for part of the relative vertical movement of the spot, the latter need not traverse the entire height h, so that the height h' of the raster in the film plane will also have to be smaller than h. (The film plane, i.e. the plane in which the film moves during the scan, should be clearly distinguished from the film itself: the film is in motion whereas the film plane is stationary.) The height h' can easily be calculated. Since v, T_s and h are known, the value of w follows from (1) and h' from

$$h' = wT_s . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

Since $h' < h$, the ratio of height to width of the raster scanned by the spot will be smaller than the 3 : 4 prescribed for a normal television raster. The same obviously applies to the raster traced by the source on the scanning tube.

The distance s between two film frames (the frame pitch) being 19 mm, the film velocity v is equal to 25 × 19 mm/sec. The standard height and width of the part of the film frame to be projected are 15.2 and 21 mm respectively, and their ratio is therefore 3 : 4.14. The ratio needed for a television picture is 3 : 4, so that to avoid showing on the television screen a part of the film frame not intended for projection, the whole height but not the whole width of the film image should be scanned. For h in (1) we may therefore insert 15.2 mm, and since $T_s = 18.5 \times 10^{-3}$ sec, we can now calculate w. The height of the raster in the film plane then follows from (2) as $h' = 6.4$ mm. The width must be $^4/_3 \times 15.2 = 20.3$, which means that the ratio of height to width is not 3 : 4 but 1.26 : 4.

Interlaced scanning

Since the film moves at a constant speed, it will have travelled half the frame-pitch during the time $T_{s1} + T_{b1}$ (20 msec) that elapses between the start of an odd raster and the start of an even one. In the film plane, therefore, the envelope of the even raster should not coincide with that of the odd raster but should be displaced over half the frame pitch ($s/2 = 9.5$ mm) in the direction of film travel. If this is done the lines of the two rasters will be correctly interlaced on the film.

The requisite displacement of the even raster is effected in the present equipment by projecting a second image of the light source on to the film by means of a second objectieve mounted below the first, in such a way that the distance between the two images is equal to exactly half the frame pitch (*fig. 3a*). A rotating shutter, directly behind the film, intercepts the beam from the lower objective during the periods T_{s1} and that from the upper objective during the periods T_{s2}. The requirements as regards the height and speed of the shutter blades can be seen from *fig. 4*, which shows as a function of time the vertical deflections of the upper and lower spots as well as the vertical positions of the upper and lower edges of a succession of shutter blades. The regions between these latter pairs of lines are hatched; the distance between them, measured perpendicular to the abscissa, represents, of course, the height of the blades. It is assumed that the blades are moving vertically upwards at a constant speed, an assumption which is, of course, only approximately true, for the shutter radius is not infinitely large. The slope of the lines that represent the upper and lower edges of the shutter blades is a measure of the vertical velocity of the blades. Blade velocity and blade height should be so chosen that during all T_{s1} periods the sawtooth for the lower spot falls within the hatched area, while the sawtooth for the upper spot falls outside it. During all T_{s2} periods the reverse should be the case. It can at once be seen from the figure that it is possible to satisfy these conditions with many different combinations of blade height and blade velocity. Since it is just as important to transmit the whole of the active beam as it is to intercept the other, it is logical to make the distance between successive blades equal to the height of the blades. Fig. 4 represents the situation obtaining in the flying-spot scanner designed in this laboratory.

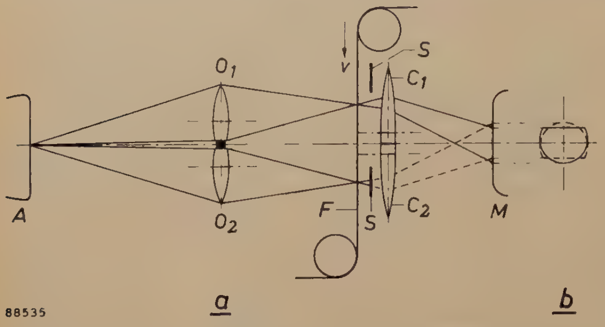

Fig. 3. *a*) To ensure that the even raster will interlace correctly with the odd, two images of the source on the scanning tube *A* are projected on to the film *F* by two identical optical systems O_1 and O_2, mounted one above the other. Two condenser lenses C_1 and C_2 collect the light passing through the film and concentrate it upon the photo-electric cell *M*. A rotary shutter *S* (not to scale) between the film and the condenser lenses alternately intercepts one of the light beams (the lower one in the figure). C_1 and C_2 project the images of O_1 and O_2 respectively upon the light-sensitive layer of the photo-electric cell M and thereby spread the light from the spots over the whole images.
b) The images of O_1 and O_2 projected on the photo-electric cell by C_1 and C_2 respectively (the image of O_2 is drawn with dashed lines). C_1 and C_2 are so spaced as to make the images coincide as closely as possible.

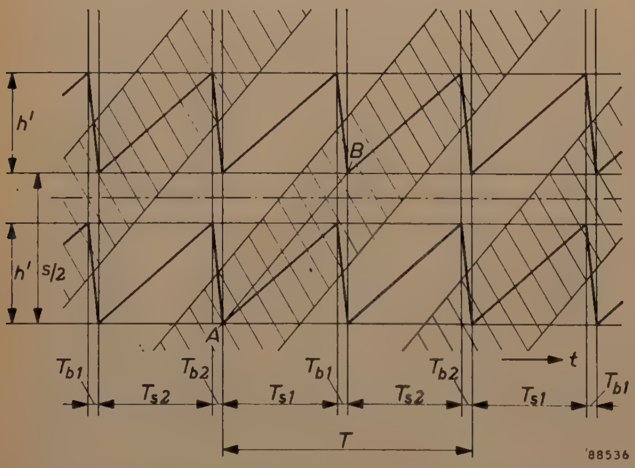

Fig. 4. Graphical representation of the alternate transmission and interception of the two scanning spots by the shutter blades. The two sawtooth lines represent the vertical deflections of the upper and lower spots as a function of time *t*. The boundaries of the hatched regions indicate the top and bottom edges of successive blades as functions of the time. In the intervals where a sawtooth falls within a hatched area the beam concerned is intercepted. Meaning of the symbols as in figs. 1 and 2.

Complications arise from the finite radius of the shutter; for example, at the inner radius the blade velocity is less than at the outer radius. However, the margin allowed is sufficient to compensate for this discrepancy, as well as for the fact that the beams of light have a larger cross-section at the position where they are intercepted than at the position of the spots themselves (fig. 3*a*).

A relation exists between the number of shutter blades *n*, the radius of the shutter (for which we take the radius r_m to the middle of the blade) and the velocity v_m of this midpoint. In $^1/_{25}$ sec the shutter must have moved exactly one blade farther, i.e. it must mave turned about an angle of $2\pi/n$ radians. The angular velocity of the shutter must therefore be $\omega = 25 \times 2\pi/n$ rad/sec. Since $v_m = \omega r_m$,

$$v_m n/r_m = 50\,\pi \quad . \; . \; . \; . \; . \; . \; . \; . \quad (3)$$

It can be seen from fig. 4 that the widest margin will be obtained if the hatched band is parallel to the line *AB*. In that case the blade velocity is equal to the film velocity (for the vertical distance between *B* and *A* is 9.5 mm, i.e. half the frame pitch, and the horizontal distance is equal to half a picture period). This value of blade velocity is accordingly taken for v_m. The value of r_m is fixed by (3) after the number of blades *n* has been decided on; *n* cannot be made arbitrarily small as otherwise r_m will also be very small and the velocities on the inner and outer edges of a blade will deviate too much from v_m. In the present case, *n* is 24, which makes r_m = 73 mm and the outer radius 85 mm.

The two light spots in the film plane are given the requisite spacing of half the frame pitch (*s*/2) by adjusting the vertical distance between the axes of the two objectives. Since there is always a certain amount of shrinkage in a film, depending upon age, humidity and temperature, the frame pitch is not precisely fixed, being invariably slightly less (about 1% max) than the nominal 19 mm. The distance between the objectives should therefore be adapted to the shrinkage in the film before beginning a transmission. Experience has shown that the amount of shrinkage in any one film does not change perceptibly over the whole length, but may vary considerably between different films. If the vertical distance between the objectives is not correctly adjusted, the lines of the two rasters on the film will no longer be properly interlaced. This does not, it is true, affect the interlace at the receiving end (where, of course, the television receiver is itself responsible for the interlacing on the picture tube) but the parts of the picture carried by the two rasters of the flying-spot scanner are no longer complementary. This is visible as a vertical doubling of the picture, a phenomenon which serves as a criterion for judging whether the distance between the objectives has been correctly adjusted.

If the line joining the centre points of the objectives is not exactly parallel with the direction in which the film is moving, the two projected images of the source will be displaced with respect to each other at right-angles to that direction. The points of a vertical line on the film, produced by the odd and even rasters respectively, will therefore appear on the picture tube displaced with respect to each other along the horizontal. This *horizontal* doubling of the picture can be avoided if the optical systems are mounted with proper care.

The central section of the film scanner is shown in the photograph in *fig. 5*.

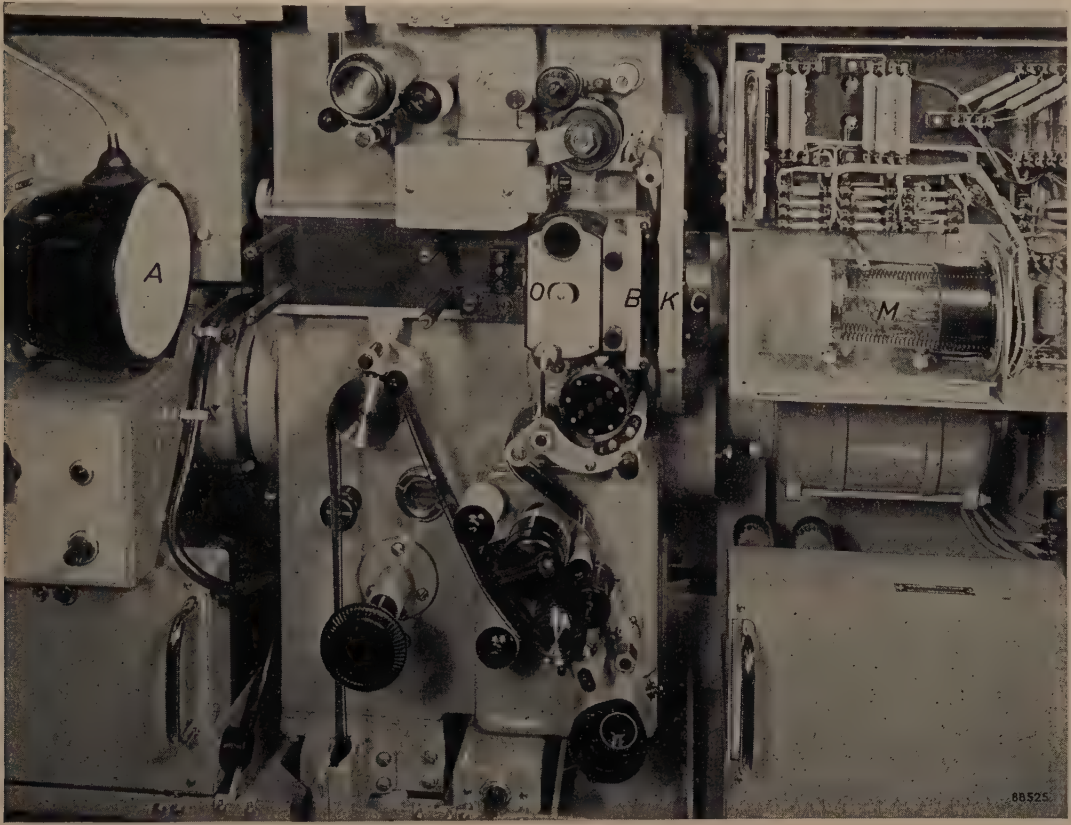

Fig. 5. The central section of the flying-spot scanner, showing the film in position. *A* scanning tube, *O* mount for double objective, *B* film gate, *K* shutter housing, *C* mount for condenser lenses, *M* photo-electric cell.

Other, perhaps more obvious, methods are conceivable for projecting the required images of the odd and even rasters in the film plane. It would be possible, for instance, to arrange for the two rasters to be traced at different positions on the scanning tube and then projected, by only one objective, in the correct relative positions on the film plane. It would then be necessary to ensure that the displaced even raster on the scanning tube is exactly complimentary to the odd raster. This is in fact impracticable, however, for with two non-centrally disposed rasters, it presupposes the strictest accuracy in the deflection of the source on the scanning tube. The method used in the present equipment imposes the requirement of strict accuracy upon the objectives, which are more readily controllable.

The objectives

Equality of focal lengths

The focal lengths of the two objectives may differ only very slightly. For convenience we shall regard the two objectives as consisting of a single lens each. When two lenses with different focal lengths have a common object (i.e. common object plane) they must be situated at different points along the axis if they are both to project a sharp image on a common image plane (*fig. 6*). The object is then projected with different degrees of magnification (the formula for magnification being $m = b/v$, where b is the image distance and v the object distance). One consequence of the difference in magnification is that the lines of the odd and even rasters on the film are not equal in length. Thus, if a vertical line in the middle of a film frame is correctly reproduced on the television picture tube, vertical lines at some distance from the middle will show the horizontal image-doubling discussed in the preceding section. Moreover, the two rasters will not be correctly interlaced over the whole of the film frame. If the interlace is correct in the middle of the frame, it will deviate above and below, resulting in vertical image-doubling at the top and bottom of the picture tube.

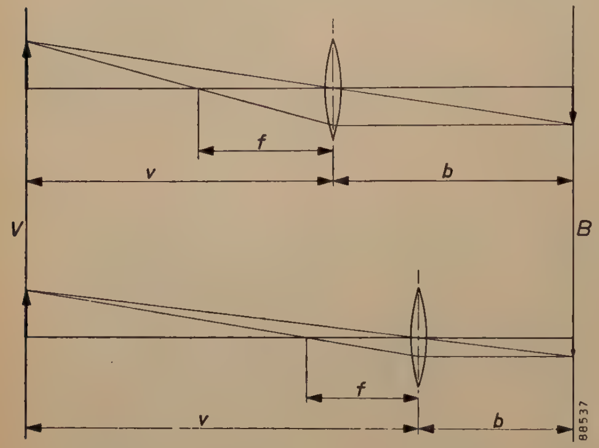

Fig. 6. Two lenses of different focal length, but with a common object plane (V), must be situated at different positions in order that both project a sharp image on a common image plane (B). The magnifications given by the lenses then differ. v = object distance, b = image distance, f = focal length.

Since the width of a television picture is greater than its height, and image-doubling increases with the distance from the middle (assuming that the picture is well adjusted in the middle) horizontal image-doubling is the more serious fault. In this connection we have stipulated that at the vertical edges of the picture the reproductions of the same line by the odd and even rasters should not be relatively displaced by more than $^1/_4$ of the line spacing. (The line spacing is the vertical distance between two adjacent lines of the odd and even rasters.) This means that the widths of the two rasters on the film may differ by at most half the line spacing on the film (viz. $^1/_4$ at the left and $^1/_4$ at the right extremity). The width of a television picture being about $^4/_3 \times 625 \approx 800$ line spacings, the magnifications caused by the two objectives should not differ by more than about $^1/_{1600} \approx 0.06\%$. Given this condition, a simple calculation shows that the difference between the focal lengths of the optical systems should be less than about 0.04%.

From the lens formula $1/b + 1/v = /1f$, the magnification formula $m = b/v$ and the stipulation $b + v = \text{constant} \doteq c$ (see fig. 6) a relation follows between the focal length f and the magnification m:

$$f = c\frac{m}{(1+m)^2}.$$

Taking first the logarithm, and then differentiating, we find:

$$\frac{\mathrm{d}f}{f} = \frac{1-m}{1+m}\frac{\mathrm{d}m}{m}. \quad \ldots \ldots \ldots \quad (4)$$

We have already seen that the permissible relative variation in magnification is 0.06%, so that $\mathrm{d}m/m = 0.6 \times 10^{-3}$. In the present case $m \approx {}^1/_5$. By inserting this value and that for $\mathrm{d}m/m$ in (4) we find the permissible relative difference between the focal lengths to be:

$$\frac{\mathrm{d}f}{f} = 0.4 \times 10^{-3} = 0.04\%.$$

The objectives used are made up of several components. Owing to the tolerances in the spacing between the components, there is often a variation of several per cent in the focal lengths, even in objectives from the same production series. By very slightly modifying the spacing of the components in one of the objectives it is possible nevertheless to make the two focal lengths equal to within the prescribed accuracy of 0.04%. This alters the lens correction but does not worsen it. The method of testing the equality of the focal lengths, in itself quite a problem, cannot be discussed within the scope of this article.

Vignetting

Another requirement to be satisfied by the objectives is that the illumination of the image point should decrease only very slightly with angular deviation from the optical axis, i.e. when the corresponding object point is moved further away from the optical axis. The scanning lines *1* and *2* in *fig. 7* will make this clear. On the scanning tube these lines are next to each other. Their active images, however, are produced by different optical systems, i.e. by the upper and lower objectives respectively. The two beams of light make different angles with the optical axes of the objectives and therefore the two images receive different quantities of light.

On the picture tube the lines again appear one above the other, but with alternating luminous intensity. As a result, the picture flickers at a frequency of 25 c/s, particularly at the upper and lower edges, where the differences are most pronounced. In the middle of the picture the odd and even scanning lines receive equal quantities of light.

Since the eye is highly sensitive to flicker of such a low frequency [6]), the maximum difference between the quantities of light received by the different scanning lines should not amount to more than a few per cent. In every lens the intensity of light in an image point diminishes as the angle α of the incident light increases, for the light at the image point is proportional to $\cos^4\alpha$ (the $\cos^4$ law). If the lenses give rise to vignetting, however, the phenomenon is much more serious. By vignetting of a lens is meant the increasing interception of the light beam, by *multiple* rims or diaphragms, with increasing α. The objectives used in the present flying-spot scanner have been so designed as to cause no vignetting at the values of α concerned, thus keeping flicker within permissible bounds.

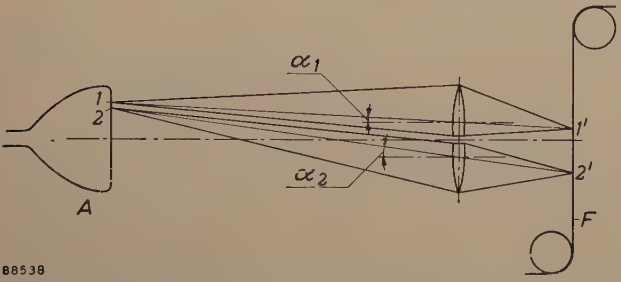

Fig. 7. The light in an image point becomes weaker the further the object point is removed from the optical axis. The image 2′ of the raster line 2 projected by the lower objective will therefore receive less light than image 1′ of raster line 1 projected by the upper objective. If this difference is not kept within certain limits it becomes noticeable as a 25 c/s flicker in the final picture.

Further particulars of the objectives

It is also necessary to ensure that the objectives cause no distortion. Slight distortion would not be so serious if both rasters were distorted in the same way, so that the lines would still interlace properly on the film. This is not the case, however; even very slight distortion causes some loss of interlace.

As discussed, the images projected by the objectives must be about 9.5 mm apart, which means that the distance between the axes of the lenses must be somewhat less than 9.5 mm (fig. 3*a*) . To make this possible, it was nesessary to saw off segments of the objectives, as illustrated in *fig. 8* [7]).

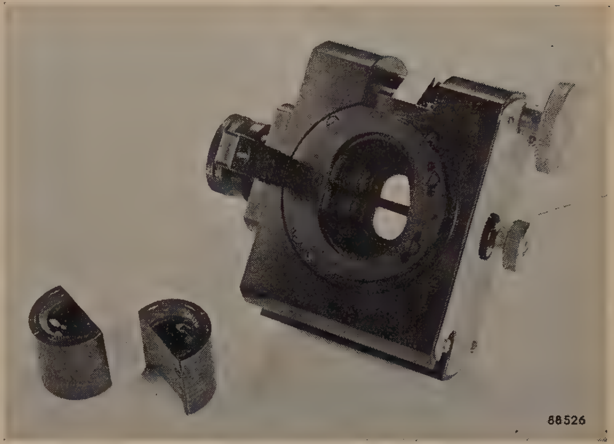

Fig. 8. The lens mount for the double objective. The small central knob on the right is used for changing the spacing between the objectives so as to obtain correct interlacing. Segments have been sawn off both objectives systems as shown.

The condenser lenses

Behind the film are mounted two condenser lenses which collect all the light passing through the film from the two rasters and concentrate it upon the photo-cathode of a photo-electric cell. Each lens projects the image of its corresponding objective upon the photo-cathode (thus the spot moving over the film does not appear as a sharp image on the photo-cathode). Consider, for example, the light beam from the upper optical system (fig. 3*a*). Irrespective of the position of the source on the scanning tube, a ray of light passes through each point of the objective (which is the object for its corresponding condenser lens) which is then transmitted to the corresponding image point on the photo-cathode. The light beam is thus continuously spread over the entire image of the objective on the photo-cathode, so that the effects of local variations of sensitivity in the latter are eliminated. Moreover, by correctly adjusting the distance between the axes of the condenser lenses, the images which they project of the corresponding objectives can be made approximately to coincide (fig. 3*b*). (They can never coincide exactly because of segments having been sawn off the objectives). As a result, the photo-electric cell is equally sensitive for both rasters.

---

[6]) J. Haantjes and F. W. de Vrijer, Flicker in television pictures, Philips tech. Rev. **13**, 55-60, 1951/52.

[7]) For the selection of the most suitable make and type of objective, the adjustment of their components and the removal of the segments, we are indebted to the optical department of Philips Research Laboratories, under the direction of P. M. van Alphen.

Slip caused by film shrinkage

The method of film scanning discussed in the foregoing is based on the assumption that the film moves at a constant speed of 25 frames per sec. Where there is shrinkage in a film the speed of transport should be adjusted in order to maintain this value. In fact, however, no provisions are made for matching film speed to shrinkage. The film is driven via a sprocket with four pairs of teeth per picture (each frame having four perforations down each side). The sprocket rotates at a constant speed corresponding to 25 frames per sec which is thus the average speed of the film. The film speed is not, however, entirely constant: if the pitch of a tooth is equal to $^1/_4$ of the nominal frame pitch, a certain amount of slip will occur four times per picture if the film is shrunk. The reason why this causes no difficulties becomes clear if we consider rather more closely what the relative positions of the two rasters in the film plane are required to be in order to obtain good interlace.

In the first place the distance between the lower extremities of both rasters in the film plane should be exactly equal to the distance travelled by the film during the period $T_{s1} + T_{b1}$. If this is not so, the first lines of the even raster on the film will not fall correctly between the lines of the odd raster. Assuming for the sake of simplicity that the film is stationary for a moment during each slip, we can plot film speed against time to obtain a curve as shown in *fig. 9*. The curve shows that during a period $T_{s1} + T_{b2}$ the film has travelled the same distance as it would have covered if its speed had the proper constant value (which is indeed equal to the average speed $\bar{v}$ in fig. 9). The distance between the two objectives must therefore be adjusted to the same value, independent of whether the film actually has a constant velocity or slips.

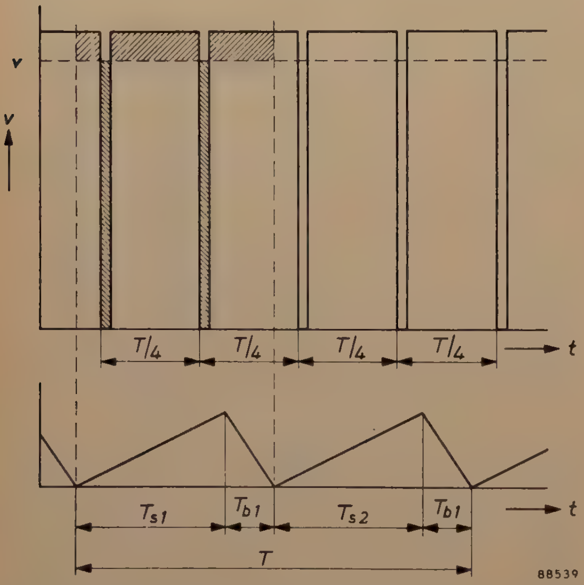

Fig. 9. Film velocity v as a function of time t when using a drive sprocket with four pairs of teeth per film frame. Slip occurs four times per picture period, so that the prescribed average speed v of 25 frames per sec is maintained irrespective of shrinkage. It is assumed that the film is stationary during slip. During an interval $T_{s1} + T_{b1}$ the hatched area above the line $v = \bar{v}$ is equal to that below the line. The lower figure represents once again the vertical deflection of the spot as a function of time.

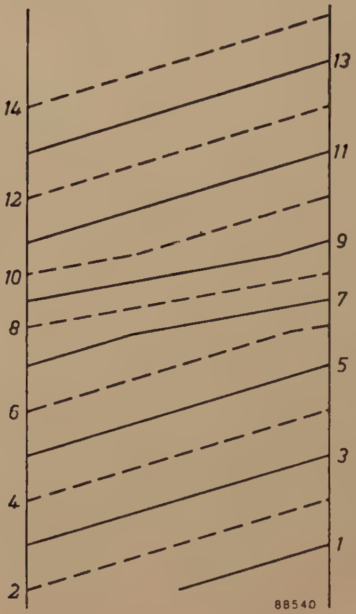

Fig. 10. Distortion of the raster on the film owing to slip. The numbers indicate the sequence of the scanning lines. There is no loss of interlace.

In the second place, the interlace — which is now correct at the lower extremities — should remain correct and not be affected by slip. If we have a film with a shrinkage of 1% (the maximum occurring in practice) the film will be stationary during each slip for $^1/_4$% of the full picture period, during which time $^1/_4$% of 625 lines, viz. $1^1/_2$ lines, will be traced. The vertical velocity of the spot with respect to the film is thus too small during the moments of slip, so that $1^1/_2$ lines of, say, the odd raster are traced with too flat a gradient (see *fig. 10*). Owing to the even number of pairs of sprocket teeth per frame and their uniform spacing, $1^1/_2$ lines of the even raster will be traced half a picture period later with the same distortion (dashed lines in fig. 10). Thus, the interlace remains correct throughout. It is clear that this would not be the case if there were, for example, three pairs of sprocket teeth per frame, that is to say an odd number of teeth. The result, with a film having maximum shrinkage (1%), is that the raster on the picture tube will be slightly distorted over 3 to 4 lines, which is not, however, perceptible. As may be seen from fig. 9, distortion of this nature generally occurs at two positions in the picture. If the film is not stationary during slip, the distortion is spread over the whole picture, but at the same time it becomes less pronounced and thus remains imperceptible in the television picture.

Summary. A flying-spot scanner for televising 35 mm film has been developed in the Philips Research Laboratories, Eindhoven. In such scanners the film moves with constant speed and this complicates the problem of interlaced scanning. The problem is solved by using two optical systems. The two objective lenses have to be closely identical and should be as free as possible from distortion and vignetting.

Since there is always some shrinkage in a film and since the drive is effected in the normal way via a sprocket, the film does not in fact travel at a strictly constant speed. A detailed explanation is given of why this is not troublesome.

ENTROPY IN SCIENCE AND TECHNOLOGY

IV. ENTROPY AND INFORMATION

by J. D. FAST and F. L. H. M. STUMPERS.

536.75:621.391

In Volume 16 of this Review one of the authors (J.D.F) devoted a series of three articles to the application of the concept of entropy to several widely varying problems in science and technology. Some of the subjects discussed were: affinity in chemical reactions, paramagnetism, the attainment of very low temperatures, heat engines, refrigerating machines and heat pumps, radiation, specific heat of solids, vacancies and diffusion in solids, alloys, the elasticity of rubber, and the solution of macro-molecules. This article, which concludes the series, is somewhat different from the previous ones in that its purport is rather negative than positive. In information theory it is possible to arrive at a concept that can be formally defined and manipulated analogously to the concept of entropy in thermodynamics. This is illustrated by various applications. However, the use of the term entropy for this new concept might wrongly suggest that all thermodynamical implications in the term were equally valid here. Some caution is therefore necessary.

Introduction

The concept of information plays an essential part in two recently developed branches of technology, viz. communication theory and cybernetics. The domain of communication theory includes the transmission of messages by means of acoustical, mechanical, electrical or optical signals. Its mathematical foundation has been largely laid down by Shannon [1]). Even more ambitious in scope is the science of control or *cybernetics*, so termed by Wiener [2]) who has done much fundamental work on this subject. The subtitle of Wiener's book: "Control and communication in the animal and the machine" states succinctly the scope of this new science. More recently [3]) the same author has particularly applied himself to the sociological aspects of his theory. At present a synthesis embracing both communication theory and cybernetics, is developing. In various symposia workers in widely diverging branches of science, such as communication engineers, linguists, psychologists and neurologists have considered to what extent their problems, which are all concerned with the transmission and the use of information, are of a common nature.

In the literature that has sprung up on this subject since about 1945, the term "entropy" is frequently encountered. Mention is made of the entropy of a source of information, the entropy of a language, the entropy of a TV picture, etc. Unfortunately, however, the word entropy is often used here in a sense that bears but little relation to the well-established thermodynamical-statistical entropy, discussed in the preceding three articles of this series [4]). We shall return to this after having first said something more about communication, cybernetics and information.

---

[1]) C. E. Shannon, A mathematical theory of communication, Bell System tech. J. **27**, 379-423 and 623-656, 1948.

[2]) N. Wiener, Cybernetics, Hermann at Cie., Paris 1947, John Wiley, New York 1948.

[3]) N. Wiener, The human use of human beings: Cybernetics and society, Houghton Mifflin, Boston 1950.

[4]) J. D. Fast, Entropy in science and technology: I. The concept of entropy, Philips tech. Rev. **16**, 258-269, 1954/55; II. Examples and applications, ibid. pp. 298-308, III. Examples and applications (cont.), ibid. pp. 312-332.

Communication

It is self-evident that communication plays a predominant role in modern civilization. No society however, primitive, is conceivable without communication — primarily in the form of a language of gestures and spoken words. In modern society this is augmented by the written and printed word, photography, telegraphy, telephony, teleprinter, radio, the cinema and television. The word communication means, quite generally, any transmission of information from one place to another. The entities and the media that participate in this constitute together a communication system. In its simplest form such a system (*fig. 1*) consists of a source of information plus a transmitter at the one end, and a receiver plus a recipient at the other end, with a communication channel linking the two.

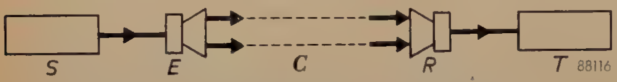

Fig. 1. Schematic representation of a communication system. *S* source of information (selector), *E* transmitter + coding device, *C* communication channel, *R* receiver + decoding device, *T* recipient.

If the communication consists of something spoken by a person *A* to a person *B*, the brain and the organs of speech of *A* constitue the source and the transmitter, whilst the ear and the brain of *B* are the receiver and the recipient. The communication channel in this case is the air.

The source of information preselects a message, i.e. it choses a specific sequence of the words, letters or figures in which the message is embodied. From this the receiver tries to reconstruct the original message, after which it is passed along to the recipient. In the case of the spoken word the signal consists of sound waves in the air, in the case of line telephony of variations in the electric current in a wire. The transmitter in the latter case is a microphone, which converts the varying sound pressure of the voice into a varying electric current; the telephone is part of the receiver. The sound waves in the air and the varying electric current in the wire correspond to the message to be transmitted according to a certain "code".

The application of carrier-wave telephony 5) makes it possible to transmit several telephone conversations over a single line, i.e. over one conductor pair. The low-frequency electric vibrations derived from corresponding acoustic vibrations from speech, are here superimposed on the high-frequency vibrations of the carrier wave. When carriers of different frequencies are separately modulated in this way with audio-frequency speech, several conversations can be transmitted at the same time. The substantially increased efficiency of information-transmission thus brought about gives rise to the question: what is the maximum amount of information per unit time that can be transmitted over a single telephone line? Before we can answer this question we need a measure to express "amount of information". The fact that there is a finite limit to the capacity of a communication channel to carry information follows from the inevitable occurrence of thermal agitation noise (see I) in every communications system. This blurs the details of the transmitted symbols, thus limiting their intelligibility. Only after we have introduced a *measure* for information can we deal with the above-mentioned question.

Cybernetics

Cybernetics is primarily concerned with self-regulating systems and processes. A very simple example of a self-regulating system is a thermostat, with which a space is automatically kept at a nearly constant temperature. This form of control can be effected, for example, with a special type of thermometer in which the mercury closes a contact as soon as the temperature exceeds a predetermined value. The electrical circuit is arranged so that a relay is tripped which reduces the supply of heat. When subsequently the temperature drops below the desired value, the thermometer contact is broken and the relay automatically causes an increased supply of heat. The essence of the matter is that here we are not concerned with a one way sequence of cause and effect, as with a room with a normal thermometer, where the position of the mercury in the thermometer depends simply on the temperature of the room; instead, we have here a two-way sequence of cause and effect, whereby the temperature of the room is also influenced by the position of the mercury.

We may put it that the above automatic control is due to the fact that the information regarding the temperature is fed back from the thermometer to the source of heat. This *feedback* is the common characteristic of the systems with which cybernetics is concerned. The information regarding a deviation from the desired condition is fed back to the control mechanism, which solely on the strength of this information can fulfill its controlling action.

---

5) See e.g. H. N. Hansen and H. Feiner, Coaxial cable as a transmission medium for carrier telephony, (Philips tech. Rev. **14**, 141-150, 1952/53.

A more elaborate example of automatic control by information-feedback is the automatic aiming of an anti-aircraft gun. A radar installation supplies data regarding the position of an aircraft to an electronic device, which, via servomotors, moves the gun into the appropriate direction. In order to function properly and to give the right instructions to the servomotors the electronic device must not only know the position of the aircraft, but also that of the gun. Of the latter it is automatically informed by electric signals transmitted by a device connected to the gun. These signals thus again provide the necessary feedback of information.

Control instruments of the latter kind are further characterized by the fact that they predict the probable future position of the aircraft (or other variable) from the data so far received. The mathematical theory of these prediction problems has been examined by Wiener and Kolmogoroff.

The ambitious scope of the cyberneticists will be better appreciated when one reflects on the many automatic processes that are continuously occurring in living creatures e.g. in the human body: its breathing, its heartbeat, its reaction to stimuli, etc. Many of these processes are involuntary, being controlled by reflexes, which are formally comparable to the above-mentioned mechanisms employing the feedback principle. An example is the flexor-reflex, which occurs for instance if one inadvertently touches a hot object with the hand. Along the sensory nerves a signal in the form of series of electromechanical impulses of short duration is passed to the spinal cord. Here the information is passed on to certain motor neurones, along which commands are sent to the flexor muscles of the arm in question, resulting in the withdrawal of the hand from the hot object. Regulation by means of hormones(organic substances secreted by certain glands in the body) may likewise be considered as a feedback mechanism. A well-known hormone is insulin, which is produced in the pancreas and plays an important part in regulating the metabolism of carbohydrates. We are probably justified in saying that all vegetative functions in the body, e.g. the control of the composition, the pressure and the pH of the blood, the osmotic pressure, the body temperature, and the composition of the gas in the lungs, are based on a form of feedback, either hormonal or nervous in character. This control has the task of providing the tissues with a constant environment, which is indispensable to the proper functioning of the organism. The name "homeostasis" has been given by Cannon [6]) to this system of regulation.

With any automatic control by feedback alone we have to put up with minor deviations from the desired value, since it is precisely these deviations that are employed to execute the necessary corrections. We are thus always concerned with oscillations around the desired value and it is our aim to reduce these to a minimum. Some cyberneticists believe that it will eventually become possible to mitigate the recurring economic crises of our society by the application of the methods of feedback engineering. They proceed from the premise that these economic fluctuations are based on a feedback of information in the sense that economic activity is determined by the rate of investment, but that the latter, in connection with the profits to be expected, is in its turn dependent upon economic activity [7]). They hold that by taking certain timely action it should in principle be possible to put a curb on economic fluctuations.

A field of great importance is opened up by the application of the principles of cybernetics to the development of "automatic factories", that is the gradual but increasing change to almost wholly automatic plants and machines.

With this "automation" the human element is entirely or partially replaced by robots that act in a controlling and regulating capacity upon the manufacturing process. Like the thermometer in the thermostat, the robot performs both the measuring and the regulating functions. It may be given the power of human muscles, or more than human power by amplification, and may even be given, in the form of an electronic computer, some of the elementary faculties of the human mind. Modern electronic computers are capable of completely taking over certain secondary functions of the human brain and within these restricted fields of action often improve on the brain with regard to speed and accuracy.

The concept of information

In daily life we speak of receiving information when we learn something that was yet unknown to us. Information removes uncertainty, or rather, *may* remove uncertainty. For the communications engineer only the latter is of importance. He is interested only in how much information can be contained in a message, and how much difficulty is involved in transmitting it. Its significance to the receiver or its intrinsic value do not concern him.

---

[6]) W. B. Cannon, The wisdom of the body, Paul, Trench, Trubner & Co., London 1939.

[7]) A. Tustin, The mechanism of economic systems: An approach to the problem of economic stabilisation from the point of view of control-system engineering, Heinemann, London 1953.

For the communications engineer all messages of equal length (i.e. symbol content) are equivalent.

For the person who wishes to transmit a message the choice of the symbols is of considerable importance. If the message to be transmitted consists, e.g. of a series of 1000 symbols, chosen from an alphabet of 32 symbols (26 letters + some punctuation symbols) then the message will be one out of a total of 32^{1000} possible messages. If the message were to contain only one symbol, there would be 32 possibilities; if the message were 2 symbols long, then each of the 32 symbols might be followed by any of 32 symbols, so that there would be 32^2 possibilities, etc. In normal practice, as regards conversation, telephony and telegraphy, one is restricted to those series of symbols that constitute an intelligible message in one of the known languages or in an accepted code. For a message of 1000 symbols one then still has the choice of a multitude of possibilities, but considerably less than 32^{1000}.

Let us, for the time being, consider the symbols as independent of each other and occurring with equal probability; then the general expression for a number q of possible messages of length N symbols, selected from m different symbols is $q = m^N$. Let another message have a length of M symbols, then the number of possible selections is $q' = m^M$. If the two messages are transmitted in direct succession, and are considered as a single message, then a choice has been made out of $Q = m^N \times m^M$ possible messages. On the other hand it is only logical to claim that the total service rendered by the telegraph office is the sum of the two separate services. These services consist of the transmission of information. We must, therefore, find a definition of the concept "quantity of information" I as a function of the number of possible messages, in such a form that:

$$I\,(m^N) + I\,(m^M) = I\,(m^{N+M})\,.$$

To express the additive character of the function, the quantity of information in a message of a given length is defined as the logarithm of the number of possible messages of that length. For the case considered here we can put:

$$\begin{array}{ll} \text{First quantity of information} & : \log q \;= N \log m \\ \text{Second quantity of information} & : \log q' = M \log m \\ \hline \text{Total quantity of information} & : \log Q = (N+M) \log m. \end{array}$$

Any separate symbol from the series of m equally probable and mutually independent symbols thus represents a quantity of information

$$i = \log m = -\log p, \qquad \text{(IV, 1)}$$

where $p = 1/m$ is the *a priori* probability of the symbols. The choice of the base of the logarithms is quite arbitrary since this merely determines the magnitude of the unit. If a choice of two equally probable possibilities is transmitted, then the quantity of information supplied is, according to (IV, 1), $i = \log 2$. This is the simplest situation conceivable, and it is therefore reasonable to choose this quantity of information as the unit: log 2 assumes the value unity if 2 is chosen as the base of the logarithms. This has in fact been generally adopted in communication theory. This unit of information is called a "bit" (short for "binary digit"). Formula (IV, 1) appropriately allots a higher numerical value to information according as the freedom of choice is greater, i.e. according as the result is more uncertain (less probable).

It will be clear from the foregoing that the use of a logarithmic function for the definition (IV, 1) is analogous to the logarithmic definition (I, 12) of statistical-thermodynamical entropy: both information content and entropy are additive, while the probabilities of independent events are multiplicative in both cases (see I, pp. 265-266).

The "entropy" of a series of events

A single symbol from the language containing $32 = 2^5$ equally probable and mutually independent symbols, supplies, according to formula (IV, 1), 5 bits of information. The symbols of a language are, in fact, not equally probable; moreover they cannot always occur independently. In order to estimate of the actual information content per symbol, we shall first of all get rid of the restriction that the symbols are equally probable. Later we shall also drop the restriction that they are independent.

Let us begin by considering an imaginary language in which the symbols occur with unequal probability and are independent of each other. For any given language the probabilities are immediately apparent from the frequencies with which the symbols occur in a number of random texts in that language. The frequency of letters in the English language is given below in Table I [8]).

It is found that letter frequencies are only slightly influenced by the subject and the author of the text.

The quantity of information conveyed by a single letter we shall define by analogy with formula

---

[8]) Taken from Fletcher Pratt, Secret and Urgent, Blue Ribbon Books, New York 1942.

Table I. Probabilities of occurrence of the various letters in the English language.

Letter	Probability	Letter	Probability
e	0.131	m	0.025
t	0.105	u	0.025
a	0.086	g	0.020
o	0.080	y	0.020
n	0.071	p	0.020
r	0.068	w	0.015
i	0.063	b	0.014
s	0.061	v	0.0092
h	0.053	k	0.0042
d	0.038	x	0.0017
l	0.034	j	0.0013
f	0.029	q	0.0012
c	0.028	z	0.00077

(IV, 1) as $-\log_2 p$, p representing the relative frequency with which this letter occurs. The question now arises how much information will be conveyed on the average by a symbol from an English text. There is a chance p_e that it will be an *e*, in which case a quantity of information $-\log_2 p_e$ will be received. There is further a chance p_t that it will be a *t*, in which case a quantity of information $-\log_2 p_t$ will be received, etc. The average amount of information per symbol received is therefore

$$H = \overline{-\log_2 p_\nu} = -\sum_\nu p_\nu \log_2 p_\nu \text{ bits per symbol.} \qquad \text{(IV, 2)}$$

With the exception of a constant we have met a formula of the same form in (I, 16), where it served as a definition of *entropy* in statistical thermodynamics. Following Shannon, the above-defined quantity H has also been designated "entropy". In the opinion of the present authors, this is a regrettable practice, not to be recommended. Instead, some such term as *information-content per symbol* is to be preferred, in view of the fact that the relationship between this information-entropy and the thermodynamical-statistical entropy is only of a formal mathematical nature, due to the common statistical background. The thermodynamical concept of entropy, however, also has physical implications: it is directly connected with the concept of temperature and with the quantity of heat that can be exchanged in the course of a process between a system and its surroundings (see I).

The definition of the (average) quantity of information per symbol, or entropy, is generally valid for series of events that can each occur with a certain probability. If only two events are possible — with probabilities p and $(1-p)$ — then the average amount of information per event is given by

$$H = -p \log_2 p - (1-p) \log_2 (1-p). \qquad \text{(IV, 3)}$$

In *fig. 2*, H is plotted as a function of p. We see that H reaches a maximum ($= \log_2 2$ or 1 bit) when the two possibilities are equally probable, i.e. when $p = 1 - p = 0.5$, in other words, when the *a priori* uncertainty with regard to the result is greatest. In the extreme cases, $p = 1$ (certainty) and $p = 0$ (impossibility), formula (IV, 3) gives $H = 0$, since then there exists no uncertainty whatsoever. Returning from formula (IV, 3) to the general formula (IV, 2), we observe that in this case too the quantity of information per event (i.e. H) is a maximum if all probabilities are equally great, which means that there is a maximum freedom of choice, that is to say, there is the greatest uncertainty with regard to the outcome.

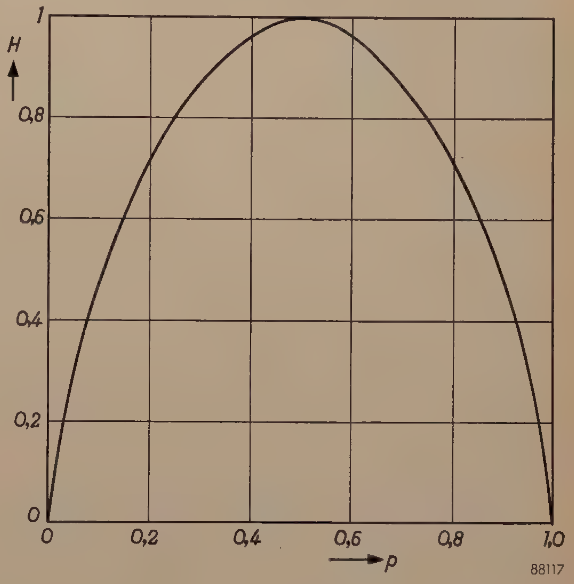

Fig. 2. The function $H = -p \log_2 p - (1-p) \log_2 (1-p)$.

In connection with the above considerations, J. F. Schouten [9]) has pointed out that the activity or "manipulation" leading to the information should be distinguished from the information itself, the latter being merely passive. Schouten expresses the quantity of manipulation in *bics* (binary choice); it is equal to the maximum quantity of information (in bits) that one may *expect* from a process. This distinction between the "active" unit bic and the "passive" unit bit has not been generally recognized.

As an illustration we may mention the well-known radio parlour-game "Twenty Questions" in which the team has to find the name of an object by putting questions that can be answered only by yes or no. The best policy for the team is to put questions that have an equal chance of being answered by yes or by no. A simple example to illustrate this principle is the following. If an object is to be located on a chessboard ($64 = 2^6$ squares), then the quickest procedure is to divide the board each time into halves (lower half, upper half, left half, and right half, etc.). In this way only 6 questions are necessary. Somewhat more complicated is the following problem. We have a pair of scales (without weights) and twelve coins of identical appearance. *One* of the coins is counterfeit. The question is how many weighings are at least required to decide which coin is counterfeit and whether this coin is too light or too heavy. We must, therefore, locate one coin out of twelve and furthermore make a choice of two possibilities. This requires $\log_2 12 + \log_2 2 = 4.585$ bits of information. One weighing may have 3 results: equilibrium, the scale tips to the left, the scale tips to the right. We can, therefore, expect $\log_2 3 = 1.585$ bits of information at the most from each weighing. For this reason it may be expected that 3 weighings will be sufficient provided that they are carried out properly. If we arrange our experiments in such a way that the 3 possible results of every weighing are about equally probable, this can indeed be realized, though not without some trouble.

---

[9]) J. F. Schouten, Proceedings of the Symposium on Information Theory, London 1950, page 195.

The total amount of information required ($\log_2 24$ bits) is less than the maximum amount of information that *can* be derived from three weighings ($3 \times \log_2 3 = \log_2 27$ bits). Consequently to solve the problem in three weighings a number of slightly different strategies may be possible. An example of a good strategy is the following.

Divide the twelve coins into three groups of 4 and weigh two groups against each other. If the scale dips, then we know that one of these groups of 4 may contain a coin that is too light or the other group of 4 may contain a coin that is too heavy. For the second weighing we place on *each* scale pan two of the coins that are possibly too light together with one of the coins that are possibly too heavy. If there is equilibrium then we know that one of the 2 remaining coins is counterfeit and too heavy. A single weighing will then show which of them is counterfeit. If there is no equilibrium then we weigh the two possibly-too-light coins that were on the scale-pan that went up, against each other. If with this third weighing the scale dips, then the coin on the scale-pan that goes up is counterfeit and too light. If there is equilibrium however then the coin that was possibly too heavy on the scale-pan that went down in the second weighing is the culprit.

There remains the possibility that in the first weighing the scales were balanced. We than take three of the four remaining coins and one of the 8 good coins and place 2 of these four coins on each scale. If there is equilibrium then the remaining twelfth coin is counterfeit and a single weighing will establish whether it is too light or too heavy. If there is no equilibrium, then we weigh the two coins on the pan not containing the good coin against each other. The result of this third weighing is then decisive. It is instructive to evaluate the information obtained from these experiments by means of formula (IV, 2) for each weighing. A similar puzzle concerns 13 coins of identical appearance, one of which *may* be counterfeit. Here the problem is to find the counterfeit coin (if there at all) and to decide whether it is too light or too heavy in three weighings. A 14th, good standard coin is provided. In this problem both the required and the maximum obtainable quantity of information are $\log_2 27$ bits, and there exists only one good strategy.

The "entropy" of a language

If we evaluate the average quantity of information per letter of the English language from the probabilities given in Table I, then we obtain 4.16 bits/letter. If the 26 letters had had equal probabilities, then H would have been $\log_2 26 = 4.70$ bits/letter. The quantity of information per letter has therefore been reduced [10]), even for an independent choice. Moreover, the probability of a letter is also somewhat dependent on the letters earlier transmitted or written down. We shall illustrate this with 2 examples: 1) The probability of the letter *u* is not particularly great, but the probability that a *u* follows a *q* approaches unity. 2) If we have received, as part of a message, the following symbols: Arrived Saturday mor:..." then it is highly probable that the next letters will be "ning", although "ose", "e", "eover" or "occo" are not inconceivable. Many other examples of correlation could be given. They all show that the number of different meaningful messages which are possible in a message of a given length is reduced by the presence of correlation.

An artifice to compute roughly the quantity of information per symbol is the following: let p_ν in formula (IV, 2) represent the probability of the occurrence of super-symbols each of length A symbols. According as we make A larger, the inter-correlations between the super-symbols will become smaller. The super-symbols, if sufficiently long, can therefore be chosen nearly independently of each other. In a very long message of N super-symbols there will be on the average $p_1 N$ specimens of the first super-symbol, $p_2 N$ specimens of the second super-symbol, etc. For the information per super-symbol we then obtain, according to (IV, 2):

$$-\sum_\nu p_\nu(A) \log p_\nu(A).$$

---

[10]) When determining the probability per symbol we should actually also take the spacings and punctuations into account. This has not been done here, in order to make use of existing tables.

Since each super-symbol comprises A symbols, the average information per symbol is represented by:

$$H \doteq -\lim_{A\to\infty} \frac{1}{A} \sum_{\nu} p_{\nu}(A) \log p_{\nu}(A) \text{ bits per symbol.} \quad \ldots \quad \text{(IV, 4)}$$

An increasingly closer approximation to H can be reached by choosing successively 1, 2, 3, …… for A, i.e. by determining the frequencies with which single symbols, symbol pairs, symbol triples, etc. occur in the language. Another estimate of the quantity of information per symbol can be arrived at by establishing experimentally how many times a person has to guess before he has found the next letter in a given text, and then the next letter, etc. There have not yet been enough experiments of this kind made to establish accurately the actual information content per symbol of a language, but it has been found that for the modern European languages the information content is around 1.5 bits per symbol.

The part played by the statistical structure in a language has been aptly demonstrated by Shannon [1]). He first considers a sequence of symbols constructed from a 27-letter alphabet (26 letters and a space) each symbol being chosen quite randomly (i.e. with equal probability) and independently. This can be regarded as a zero-order approximation to any language using the roman alphabet. The first-order approximation to (say) English is then obtained by choosing successive letters independently but with each letter and the word spacing occurring with the natural frequency that they have in English (see Table I). Thus *e* is chosen with probability 0.13 (roughly its relative frequency in English) and *w* with probability 0.015. For the second order approximation, pair-structure [11]) is introduced: after a letter is chosen, the next one is chosen in accordance with the natural frequencies of the various pairs (e.g. th, ed, are frequent pairs in English). In the third-order approximation, triple-structure is introduced: each letter is then chosen with probabilities which depend on the preceding two letters. Proceeding in this way we can construct better and better approximations to English:

a) Zero-order approximation (symbols independent and all equally probable).
xfoml rxkhrjffjuj zlpwcfwkcyj ffjevvkcqsgxyd qpaamkbzaacibzlhjqd

b) First-order approximation (symbols independent but with frequencies as in English).
ocro hli rgwr nmielwis eu ll nbnesebya th eei alhenhttpa oobttva nah brl

c) Second-order approximation (pair-structure as in English)
on ie antsoutinys are t inctore st be s deamy achin d ilonasive tucoowe at teasonare fuso tizin andy tobe seace ctisbe

d) Third-order approximation (triple-structure as in English).
in no ist lat whey cratict froure birs grocid pondenome of demonstures of the reptagin is regoactiona of cre

In the last example (3^{rd}-order approximation) some words and parts of words are beginning to be recognizable. Further approximations become more and more time-consuming and are therefore not attempted.

An analogous synthesis, and one that sooner resembles the language, may be achieved by considering whole words instead of letters:

e) First-order word approximation (words chosen independently but with their natural frequencies).
representing and speedily is an good apt or come can different natural here he the a in came the to of to expert gray some to furnishes the line message had be these.

f) Second-order word approximation (the transition probabilities of the words, i.e. their natural pair frequency is taken into account).
the head and in frontal attack on an english writer that the character of this point is therefore another method for the letters that the time of who ever told the problem for an unexpected.

Again the resemblance to normal English increases noticeably at each stage.

All these examples clearly demonstrate that the more we take the structure of the language into account the more the choice is restricted. The information content per symbol is consequently smaller than it would be with a free, independent choice. This difference between the maximum information content per symbol and the actual information content is called the *redundancy* R:

$$R = H_{\max} - H. \quad \ldots\ldots \quad \text{(IV, 5)}$$

---

[11]) Pair and triple frequencies are also given in the book by Fletcher Pratt quoted in footnote [8]). Word frequencies are tabulated in Relative frequency of English speech sounds, by G. Dewey, Harvard University Press 1923.
Shannon [1]) also mentions a simple method for constructing these examples without the use of such tables: To construct a second order letter approximation, for example, open a book at random and select a letter at random on the page. This letter is recorded. The book is then opened to another page and one reads until this letter is encountered. The succeeding letter is then recorded. Turning to another page this second letter is searched for and the succeeding letter is recorded; ond so on. This process was actually used by Shannon for constructing the examples (*c*) and (*d*) and for the word approximations (*e*) and (*f*).

The *relative redundancy* is also often considered; this is the ratio of the redundancy to the maximum information content:

$$r = \frac{H_{\text{max}} - H}{H_{\text{max}}} \quad . \quad . \quad . \quad . \quad . \quad \text{(IV, 6)}$$

For languages such as English, French, German, Dutch the relative redundancy is 60-70%. (At the beginning of this section, p. 206, H_{max} is evaluated at 4.70 for the roman alphabet, whilst on the next page, H is given as 1.5 for the modern European languages.) For the telegraphic transmission of information in one of the above languages it would in principle be possible to save 70% of the time by converting the messages into such a code as to reduce the redundancy almost to zero. This obviously requires the coding *en bloc* of long series of symbols which have hardly any mutual correlation left. The use of such a complicated code would in its turn cause a certain delay. This might be acceptable if it were not for the fact that a finite redundancy has some advantages of its own. A finite redundancy makes it possible to reconstruct a message mutilated by interference or noise at the receiving end. If we receive a telegram: "Adrive mext suntay" then we assume immediately that the meaning was "arrive next sunday". In a code with maximum information content per symbol, however, any alteration will completely alter the significance of the message and it is by definition no longer possible to recover the original meaning from the context. Since distortion during transmission is inevitable, irrespective of the system of communication, (noise with electric transmission and ambient noise with the spoken word), a language should always have a certain redundancy. The fact that the redundancies of the known languages, in spite of (or because of?) their very long history, vary only slightly gives rise to the supposition that this "general" redundancy is not far removed from the optimum value [12]). A slight reduction of the redundancy, however, is possible without harm being done. This is demonstrated by certain of the simplifications adopted in the American version of the English language.

The "entropy" of a television signal

The general observations on the "entropy" of a series of events are applicable to every form of communication. We shall deal here briefly with their bearing on television. Suppose that the separate elements of an ordinary black-and-white television-picture can assume a number n of different brightness levels. Let us again indicate the probabilities of the occurrence of these levels by p_ν, then the "entropy" of every television picture could be expressed by

$$H = -\sum_{\nu=1}^{n} p_\nu \log p_\nu \text{ bits per picture element,}$$

if the brightness levels of the separate picture elements were completely independent of each other. (Here, too, it would be better not to speak of entropy but of the quantity of information per picture element.) The brightness levels of the picture elements, however, show a definite correlation with regard to both place and time. Most picture elements in an ordinary picture vary only slightly in brightness from adjacent elements. We would furthermore not get an impression of continuous motion if the majority of the picture elements did not vary gradually in brightness with time. Like a language, the television picture shows a considerable redundancy, and similarly, a picture that only possessed the statistically correct brightness distribution would never approximate to a normal picture. The efficiency of the transmission of information from the television transmitter to the receiver is decidedly impaired by this considerable redundancy. This is manifested by the necessity of a very large bandwidth for the present television systems. Recently some interesting suggestions were made as to how, with the aid of a suitable system of coding, the redundancy and hence the bandwidth of the signal to be transmitted might be reduced [13]). Information theory defines the limits of what is attainable, but the future must decide how closely our technology can approach them.

Capacity of a communication channel with noise

In the foregoing we have mentioned in passing the possible interference with and mutilation of the information content of a message during transmission. As a rule the signals comprising the message in some form of code reach their destination via a transmission path (telephone line, radio link) in which interference occurs. The main source of interference, which is universally present, is random noise. This is caused by the thermal agitation in resistors and valves of transmitter, receiver and amplifying stations, and by many other random fluctuations on the transmission path. Owing to the interfering influence of the noise, we cannot be certain whether the signal received is the one that was actually transmitted. The information received is consequently less than it would be if no noise were present. Let us consider a case in which there are a total of n symbols that can all be transmitted with equal *a priori* probability. The reception of each symbol would then supply, in the absence of noise, $\log_2 n$ bits of information. If, however, the presence of noise has the effect that reception of a given symbol only means that we have a choice of three equally probable "adjacent" symbols, then we lack $\log_2 3$ bits of information, and we have thus received only $\log_2 n - \log_2 3$ bits.

---

[12]) Exhaustive studies on this and other structural phenomena of the language have been made by B. Mandelbrot. See e.g. Structure formelle des textes et communication, Word **10**, 1-17, 1945.

[13]) E. C. Cherry and G. G. Gouriet, Proc. Instn. Electr. Engrs **100**, 9-18, 1953.

A similar influence of the noise on a communication system, can be imagined as follows: the information contained in a message is coded at the transmitter end in the form of short, similar electric pulses, which are transmitted in rapid succession (*fig. 3*). Each letter then corresponds, say, to a specific pulse height. The effect of noise will now be that the pulses arrive at the receiving end in a mutilated condition, so that there will be some degree of uncertainty regarding the original height of each pulse. According as the influence of the noise is greater compared with the minimum difference in level between two pulses representing different symbols, so the uncertainty regarding each received signal will be greater and the information obtained correspondingly less.

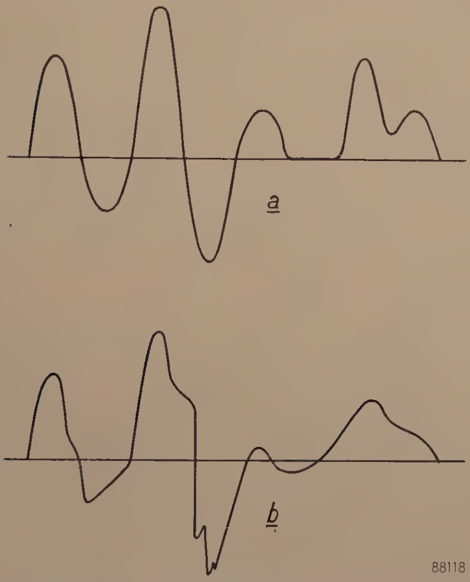

Fig. 3. Transmission of information by series of sinusoidal pulses. The information is contained in the amplitude of the pulses. At the receiving end (*b*) the pulses are received in a mutilated condition, so that an unequivocal recognition of the original symbol at the transmitting end (*a*) is not possible.

It is possible to formulate these thoughts in a somewhat stricter mathematical form. In the following, some general and well-known formulae are derived concerning the maximum quatity of information per symbol that can be obtained at the receiving end of a communication system with noise, and concerning the minimum power that is required for the transmission of information. Remarkably enough, with this latter problem, the thermodynamical entropy re-enters the picture.

Let the probability that a given symbol is transmitted be p_ν. In the case of a correct transmission of mutually independent symbols, the average information content per symbol at the receiving end would amount to

$$H = -\sum_\nu p_\nu \log p_\nu \quad . \; . \; . \; . \quad \text{(IV, 2)}$$

This and all subsequent summations have to be made over the whole range of available symbols.

Let $p_{\mu\nu}$ be the chance that a transmitted symbol μ is detected at the receiving end as the symbol ν. $p_{\mu\nu}$ may be called a *transition probability*. In a noise-free channel

$$p_{\mu\nu} = \begin{cases} 1 \text{ if } \mu = \nu. \\ 0 \text{ if } \mu \neq \nu. \end{cases} \quad . \; . \; . \; . \quad \text{(IV, 7)}$$

If a random, long series of symbols is transmitted, then the nett chance that each received symbol will be ν is

$$q_\nu = \sum_\mu p_{\mu\nu}\, p_\mu \quad . \; . \; . \; . \quad \text{(IV, 8)}$$

and the chance that any received symbol ν is actually the result of the transmission of a symbol λ is

$$P_{\lambda\nu} = \frac{p_\lambda p_{\lambda\nu}}{q_\nu} = \frac{p_\lambda p_{\lambda\nu}}{\sum\limits_\mu p_\mu p_{\mu\nu}} \quad . \; . \; . \quad \text{(IV, 9)}$$

The reception of symbol ν leaves, according to (IV, 9), an additional chance for each symbol that it was *actually* transmitted as such. The uncertainty as to whether the received symbol is the same as the transmitted symbol is therefore

$$h_\nu = -\sum_\lambda P_{\lambda\nu} \log P_{\lambda\nu} \; . \; . \; . \quad \text{(IV, 10)}$$

The *average uncertainty* after reception of any symbol is then:

$$\bar{h} = \sum_\mu q_\nu h_\nu = -\sum_\nu\sum_\lambda p_\lambda p_{\lambda\nu} \log \frac{p_\lambda p_{\lambda\nu}}{\sum\limits_\mu p_\mu p_{\mu\nu}} \; . \; . \quad \text{(IV, 11)}$$

The average quantity of information received per symbol is therefore

$$H - \bar{h} = -\sum_\nu \left[p_\nu \log p_\nu - \sum_\lambda p_\lambda p_{\lambda\nu} \log \frac{p_\lambda p_{\lambda\nu}}{\sum\limits_\mu p_\mu p_{\mu\nu}} \right] . \quad \text{(IV, 12)}$$

The noise properties of the communication channel are entirely determined by the transition probabilities $p_{\mu\nu}$. For a noise-free channel, to which (IV, 7) applies, it can be readily verified that

$\bar{h} = 0$. If the transition probabilities are given, it is possible to choose the relative probabilities p_ν of the code symbols to be transmitted in such a way that the symbols most liable to interference (with the highest h_ν-values) are least used. This means a special choice of coding. With an optimum choice of coding, $\bar{h}$ is small and the quantity of information transmitted per symbol is maximum. The quantity of information than transmitted is called the *capacity* of the communication channel. Shannon has demonstrated that the capacity of a communication channel can in principle be fully utilized by dividing each message into long series of symbols (letters) and representing each series by a unique set of code symbols (series of pulses). Of all sets of symbols that it would be possible to use, only those are chosen for the code that differ so much that they are least liable to confusion. In practice this would require a sort of dictionary of all kinds of extensive messages, ready for use and translated into sets of code symbols. The scope of the code-books and the delay in looking for the "translation" obviously impose a limit to the possibility of realizing this ideal in practice.

We shall now apply the foregoing to the earlier-described simple model of a communication system, in which coding is effected in the form of voltage pulses whose height determines the symbol. In a noise free communication channel an infinite number of symbols is in principle available, since both at the transmitting and at the receiving end a clear distinction can be made between arbitrarily small differences in amplitude. In this way any message could be expressed in a single pulse by measuring its height to millions of decimal places, the first two decimals representing the first letter, the next two decimals the second letter, etc. In reality, noise is always present, with the result that the original height of a transmitted pulse can be determined at the receiving end only with a given degree of uncertainty, so that only pulses that already possessed a finite difference in height at the transmitting end can be distinguished.

Thermal noise has, at constant temperature and in linear networks, a stationary character, i.e. the probability distribution of the fluctuating noise voltage V does not depend on time. This probability distribution is Gaussian: the probability that the voltage at a given moment lies between V and $V + \mathrm{d}V$ is given by

$$W(V)\,\mathrm{d}V = \sqrt{\frac{a}{\pi}}\,\mathrm{e}^{-aV^2}\,\mathrm{d}V. \quad \text{(IV, 13)}$$

The root mean square of the noise voltage, or the power developed in a resistance of 1Ω, then amounts to:

$$N = \overline{V^2} = \sqrt{\frac{a}{\pi}} \int_{-\infty}^{+\infty} V^2\,\mathrm{e}^{-aV^2}\mathrm{d}V = \frac{1}{2a}. \quad \text{(IV, 14)}$$

Two transmitted pulses of different height can only be distinguished at the receiving end if the difference in level, expressed as a measure of energy, is at least of the same order as the amount given by (14). A further calculation shows that if the voltage levels differ by an amount $\sqrt{2N} = \sqrt{1/a}$, there is probability of about $\frac{1}{2}$ (0.48 to be exact) that noise will cause pulse distortion to the extent of half the distance to an adjacent level. There is thus a chance of 0.48 that a transmitted pulse will be faultily received (too high or too low). The levels should therefore be placed somewhat further apart. If intervals of $\sqrt{6N}$ are used, the chance of faulty reception becomes 0.22, which is considered acceptable in practice. For a communication system of the type described here the levels will, therefore, be placed at voltage intervals of $\sqrt{6N}$. Only a finite number of pulse heights is then, of course, permissible, viz. $\sqrt{E/N} + 1 = \sqrt{s} + 1$, E representing the mean signal power, and hence s the signal-to-noise ratio (in terms of energy).

The above number is arrived at as follows. Let the maximum pulse voltage be $V_{\max}$. If the number of equidistant pulse heights is $n + 1$, than the various pulse amplitudes are given by $gV_{\max}/n$, where g may have any integral value between 0 and n. Each pulse represents a power $\frac{1}{2}g^2 V_{\max}{}^2/n^2$, and if all pulse heights occur with equal probability, the mean signal power amounts to

$$E = \tfrac{1}{2}\,\frac{V_{\max}{}^2}{n^2}\,\frac{1^2 + 2^2 + \ldots + n^2}{n+1} \approx \frac{V_{\max}{}^2}{6}.$$

Hence $V_{\max} \approx \sqrt{6E}$. Since the distance between the levels was fixed at $\sqrt{6N}$, the number of permissible levels becomes

$$\sqrt{\frac{6E}{6N}} + 1 = \sqrt{\frac{E}{N}} + 1 = \sqrt{s} + 1.$$

In practice not only are the pulse heights restricted but also the pulse frequency, i.e. the number of pulses that can be distinguished per second. This is attributable to the fact that in every communication channel only a limited frequency range is available for the transmission. Let the available bandwidth be f_0; in other words, assume that only frequencies between a given frequency f_c and the frequency $f_c + f_0$ are transmitted. It can be proved on general grounds that two signals each of duration τ seconds can no longer be distinguished if they have the same amplitude at more than $2f_0\tau$ instants.

A periodic signal of period τ can be represented by the following Fourier-series:

$$v(t) = \frac{a_0}{2} + \sum_{1}^{\infty} a_k \cos 2\pi k \frac{t}{\tau} + \sum_{1}^{\infty} b_k \sin 2\pi k \frac{t}{\tau}\,.$$

If only the frequencies between 0 and $f_0 = l/\tau$ occur, only $2l + 1 = 2f_0\tau + 1$ terms occur in these expansions, and $v(t)$ is thus completely determined by the $2l + 1$ available coefficients ($a_0, a_1, \ldots\ldots a_l, b_1, \ldots\ldots b_l$). More coefficients do not define the shape of the signal any better. The 1 may be neglected if $2f_0\tau \gg 1$ which is always the case. For a non-periodic signal of duration τ the series is replaced by a Fourier-integral, but the conclusion that $v(t)$ is completely defined by $2f_0$ data per second remains valid. These conclusions, drawn for the frequency range between 0 and f_0, are equally applicable to the arbitrary frequency range between f_c and $f_c + f_0$.

Pulse series lasting τ seconds are thus completely defined by the amplitude at $2f_0\tau$ instants. This means that the maximum quantity of information that can be contained within a bandwidth of f_0 c/s in a pulse series of duration τ seconds is attained if in that time $2f_0\tau$ pulses are transmitted.

Since, as we have shown above, a choice can be made of $\sqrt{s} + 1$ discrete heights for the transmission of each pulse, the total number of distinct (i.e. different) pulse series of length τ amounts to $(\sqrt{s} + 1)^{2f_0\tau}$. In such a series of length τ, therefore, the maximum quantity of information is

$$I = 2\,f_0\tau \log_2 (\sqrt{s}+1) \text{ bits.} \quad \text{(IV, 15)}$$

This largest possible quantity of information is what we earlier termed the channel capacity; this time, however, it is expressed in terms of the bandwidth and the signal-to-noise ratio of the communication channel in question.

The derivation given above is broadly that given by Tuller [14]). Shannon [1]) arrives at the directly comparable result:

$$I = f_0\tau \log_2 (1+s). \quad . \quad . \quad \text{(IV, 16)}$$

For strong signals this expression is virtually identical with (IV, 15). The relationship between efficiency of information-transmission and bandwidth of the communication channel, mentioned briefly above in connection with the television picture, will now be clear. The greater the redundance of the message to be transmitted the less efficiently will the channel capacity be utilized, and a bandwidth greater than follows from (IV, 16) will then be required for transmitting a given quantity of information.

Information and energy

If a signal is transmitted over a cable and we wish to receive the maximum signal energy, the cable must have a reflection-free termination. In that case, as mentioned in the first article of this series (Vol. **16**, p. 262), a noise power $N = kT_a f_0$ in a bandwidth f_0 is produced by the cable at a cable temperature T_a; k represents Boltzmann's constant.

If the signal power that can be extracted from the cable amounts to S, then the maximum quantity of information that can be transmitted in the time τ is, according to (IV, 16):

$$I = f_0\tau \log_2 \left(1 + \frac{S}{kT_a f_0}\right). \quad . \quad \text{(IV, 17)}$$

The quantity of information available per unit time and per unit signal power is then:

$$\frac{I}{S\tau} = \frac{f_0}{S} \log_2 \left(1 + \frac{S}{kT_a f_0}\right). \quad . \quad \text{(IV, 18)}$$

Considered as a function of the variable S/f_0, this function always has a negative differential coefficient Consequently it reaches its maximum in the limiting case where the signal power is zero. By expanding formula (IV, 18) in terms of $S/kT_a f_0$, the maximum available quantity of information per unit time and per unit power is found to be

$$\left(\frac{I}{S\tau}\right)_{\max} = \frac{1}{kT_a} \log_2 \mathrm{e}. \quad . \quad . \quad \text{(IV, 19)}$$

Thus, to gain 1 bit of information per second we require at the very least (i.e. for very small S/f_0) an energy of $kT_a/\log_2\mathrm{e}$, i.e. $kT_a \ln 2$. This consequence of Shannon's calculations has been pointed out by Felker [15]). According as the bandwidth is smaller and the signal power greater, the energy required per bit increases.

In this connection, communication theory provides a means of solving a classical paradox relating to the second law of thermodynamics, viz. that concerning "Maxwell's demon". Maxwell had pointed out that a being capable of acting on a molecular scale would be able to cause changes in state that are in direct conflict with the second law. Let us imagine, e.g. a gas-filled vessel whose walls are perfectly insulating. A partition, also a perfect insulator, with a small hole in it divides the vessel into two compartments. The "demon" can close this hole with a valve, which he operates in such a way that only rapid molecules are passed from right to left and slow ones from left to right. In this manner he divides the gas into a hot and a cold part, thus reducing the statistical-thermodynamical entropy of the system without doing any work. It has often been pointed out that this reasoning is unsatisfactory, for several reasons.

---

[14]) W. G. Tuller, Proc. Inst. Radio. Engrs. **37**, 468-478, 1949.

[15]) J. H. Felker, Proc. Instn. Radio Engrs. **40**, 728-729, 1953.

Brillouin in particular has drawn attention to the fact that the demon would require information regarding the speed of the molecules for his manipulations. The demon would be unable to "see" anything in a space of constant temperature and would have to use the additional energy of a "lamp". This information would cost at least as much energy as corresponds to the gain in entropy — and, in practice, far more, even with the most favourable measuring technique conceivable.

It is interesting from a historical point of view that as early as 1929 Szilard [16]), on thermodynamical considerations, came to the conclusion that 1 bit of information (though he expressed it otherwise) in a system of constant temperature would have to cost an energy of at least $kT \ln 2$ in order to avoid conflict with the second law. Szilard's imaginary experiment can be briefly described as follows. A single molecule is contained in a vessel, kept at a constant temperature by a heat reservoir. In the vessel is a piston with an opening that can be closed with a sliding valve. The piston is initially in the middle of the vessel. At the beginning of the experiment the observer closes the opening and looks to see whether the molecule is at the right or at the left. Then he allows the piston to be gradually moved in the opposite direction. The molecule hits the piston several times, thus expending work upon it. This is the energy which corresponds to the isothermal expansion of a perfect gas from the volume $\frac{1}{2}V$ to the volume V, during which an amount of heat $kT \ln 2$ per molecule is abstracted from the reservoir (This is deducible from I, equation 15). After some time the piston reaches its extreme position. The valve is now opened and the piston is returned to its initial position, after which the whole process can be repeated.

At first sight the foregoing conflicts with the second law, since energy is being continuously extracted from a constant temperature source apparently without doing any work. In order to avoid such a discrepancy, however, Szilard assumes that the measurement itself, which involves determining the place of the molecule — i.e. supplying one bit of information (right or left, i.e. two equally probable possibilities). — requires an energy of at least $kT \ln 2$. The application of ideas such as entropy, temperature, isothermal expansion etc. to this gas, which consists of one molecule, as well as the idea of action on a molecular scale in this imaginary experiment deserves to be subjected to criticism. Szilard himself partly answers this criticism by extending the argument to large numbers of molecules; others such as Raymond, have since described analogous imaginary experiments with gases consisting of more than one molecule. However, the merit of Szilard's achievement remains: he derived an interesting result which satisfactorily agrees with present ideas in information theory.

In this article we have seen that there is a connection between the "entropy" of information theory, and the statistical-thermodynamical entropy. We found this connection when determining the amount of energy needed for the transfer of information through a channel with *noise*. Since the intensity of the noise depends on the temperature, the energy must also depend on temperature, and it is therefore hardly surprising that the thermodynamical notion of entropy should come into these considerations. Discussion of Szilard's experiment, however, shows how little the term "entropy" is justified in information theory. The use of this term has sometimes been responsible for the following explanation of the result obtained by Szilard: 1 bit of "information entropy" equals a "thermodynamical entropy" of $k \ln 2$. However, the introduction of Boltzmann's constant k in pure information theory is really quite irrational (as opposed to its natural appearance in *thermodynamical* considerations of communication channels, etc.). For example, what could be the significance of Boltzmann's constant when dealing with the "entropy" of a language, where there is no question of either noise temperature or energy? What indeed could be the thermodynamical "system" of which the language forms a part? In the absence of such a system, Boltzmann's constant clearly has no place.

So long as we confine ourselves to the statistical aspect of the notion of entropy there is no danger of confusing thermodynamics and information theory. The mathematical analogy, however, does not imply that the experimental laws which are valid for thermodynamical entropy (second law), should have a corresponding physical significance in information theory.

Summary. In the relatively new branch of science known as information theory, which is mainly concerned with communication and cybernetics, the notion of information is given a quantitative meaning on the basis of statistical considerations. A quantity termed "information content per symbol" is thus derived, which is formally closely analogous to statistical-thermodynamical entropy (see the previous articles I, II and III of this series). The concept can be easily applied to any given series of events each of which may occur with a certain degree of probability, as for example the occurrence of letters in a language. This has led some to speak of the "entropy" of a language, or the "entropy" of a television picture.

---

[16]) L. Szilard, Z. Phys. **53**, 840-846, 1929.

An important application of information theory lies in the investigation of the efficiency of the transmission of information, e.g. by electric signals. By methods which show a certain formal agreement with those applied in statistical thermodynamics, it is possible to derive highly general results concerning the way in which the noise and bandwidth of the communication channel affect the capacity of the channel. Finally, the minimum quantity of energy needed for the transmission of a given quantity of information per unit time is discussed. Owing to the effect of thermal noise, it is found to depend on the temperature of the communication channel. This dependence on temperature establishes a connection with thermodynamical entropy, but the connection is based on trivial grounds, and constitutes no justification whatsoever for the common use of the term "entropy" in the theory of information. This abuse may lead us wrongly to expect analogies between thermodynamics and the information theory which go beyond their common statistical background.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk * can be obtained free of charge upon application to Philips Electrical Ltd., Century House, Shaftesbury Avenue, London W.C. 2.

2330: A. M. Kruithof: Quelques idées sur les transformations du verre (Verres et Réfractaires **9**, 311-319, 1955).

Measurements made with different glasses show that any glass can have various densities at room temperature. The difference in density is the result of different thermal histories of the glass samples. On the basis of a theory of the freezing-in of structure states, the existence of different densities at room temperature can be explained with the use of the contraction curves. Following the definition of Tool, these structure states can be characterized by a "fictive temperature". Ideas on changes in structure states can explain the mutual differences between expansion curves as well as differences between these curves and the contraction curves. The conceptions of "transformation point" and "transformation range" are discussed on the basis of this theory. Other physical properties e.g. viscosity and electrical conductivity (direct current) show similar phenomena. The curve representing the log of the specific resistance as a function of the inverse of the absolute temperature consists of three straight lines. Both points of intersection may correspond to transformations in the glass. Thermal differential analysis was used as a check on this hypothesis. That the lower point does correspond to a transformation is not, however, completely established.

2331: H. G. van Bueren: Theory of the formation of lattice defects during plastic strain (Acta Met. **3**, 519-524, 1955).

A simple theory is presented by which a relation between the plastic strain and the concentration of vacancies, interstitials and dislocations is obtained. Dislocations are formed by sources under the action of an applied stress, the other defects by the movement of jogs in the dislocations. The action of a dislocation source under a varying stress is studied, for the case of static as well as of dynamic generation. After the first few percent of strain, both methods of generation yield the same resulting defect concentrations. By suitable elimination, the influence of work hardening could be left out of the theory, and the defect concentrations could be expressed as functions of the amount of plastic strain. The theoretical deductions should be valid between plastic strains of 0.05 up to 1. They are compared with the observed critical shear stress, the elementary structure and the resistivity-strain relation in slightly deformed copper.

2332: G. Brouwer: Electrical analog of the eddy-current-limited domain-boundary motion in ferromagnetics (J. appl. Phys. **26**, 1297-1301 1955).

The observed losses in a ferromagnetic core are always greater than those calculated on the basis of a homogeneous permeability. In a magnetization process by domain-boundary displacements the permeability certainly is not homogeneous, but attains extreme values in the domain walls. As a result the loss factor due to eddy-currents is increased. This is the well-known eddy-current anomaly. The effect increases with the domain size. An electric circuit analogue was used to determine the eddy current distribution and domain-boundary motion in a number of idealized cases.

2333*: A. van Weel: Un nouveau système de mesure des angles de phase (Aéroélectronique 1er Congrès International, Dunot, Paris 1955, pp. 565-572.

If a four-pole is inserted in the feedback loop of an oscillator, the frequency of the latter is affected by the phase change suffered by the waveform on passing through the four-pole (See also Philips tech. Rev. **15**, 307-316, 1953/54.) This principle offers considerable advantages for phase angle measurement. By transforming phase variations into frequency variations, phase meters with the following properties can be designed: range 0-360° or more, precision 1°; small variations of phase measurable to an accuracy better than 0.01°; direct indication on a calibrated linear scale. For certain applications the unknown phase angle originates in a four-pole which cannot be inserted in the oscillator loop. To measure the phase relation between two A.C. voltages a new principle has been devised which permits the unknown phase angle to be introduced into a separate oscillator of special configuration. To the above-mentioned properties may then be added: insensitivity to not-too-large frequency variations in the given voltages. Possible applications include altimeters and rangefinders.

2334: H. B. G. Casimir: On the theory of superconductivity (Reprinted from: Niels Bohr and the development of physics, Pergamon Press, London 1955).

Survey of certain aspects of superconductivity and of proposed theories of this phenomenon.

2335: J. de Jonge and B. H. Bibo: The preparation of ortho and para hydroxybenzyl alkyl ethers (Rec. trav. Chim. Pays Bas **74**, 1448-1452, 1955).

Alkyl ethers of ortho and para hydroxybenzylalcohol can be obtained with a good yield by heating the hydroxybenzyl alcohol with the appropriate alkyl alcohol at 150 °C.

2336: G. Thirup: Wide-band three-phase RC-generators for complex measurements of two-poles and four-poles (J. Brit. Instn. Rad. Engrs. **15**, 597-605, 1955).

From a 3-phase R-C oscillator two voltages are derived, the complex ratio of which can be varied. The complex ratio is independent of the frequency. The two voltages are used in a composition circuit for measuring the parameters of two-poles and four-poles. Two equipments are described covering the frequency ranges 20 c/s-22 kc/s and 22 kc/s-10 Mc/s. The possible error is ± 0.5 db and ± 2° in the frequency range 100 c/s — 3 Mc/s. Outside this range the phase error may increase about 3 times while the amplitude error remains nearly the same.

2337: P. Jongenburger: The magneto-resistance of metals deformed at low temperatures (Suppl. Bull. Inst. Int. du Froid, Annexe 1955).

Carefully annealed pure copper wires were plastically deformed at 20 °K. Their magneto-resistance was measured at the same temperature before and after the deformation. When the results were plotted in a Kohler diagram, a large influence of the deformation was observed. This effect was shown to be due to dislocations only; it was independent of the simultaneous presence of point-defects.

2338: F. A. Kröger: The physical chemistry of crystal phosphors (Proc. I.R.E. **43**, 1941-1944, 1955, No. 12).

After an historical introduction, a survey is given of the present views regarding the constitution and the preparation of inorganic crystals phosphors. Particular attention is paid to the incorporation of atoms with a valency deviating from that of the atoms of the base material, and to the stabilization of atoms in a particular valency.

2339: E. W. Gorter: Some properties of ferrites in connection with their chemistry (Proc. I.R.E. **43**, 1945-1973, 1955, No. 12).

After an elementary introduction on the origin of the magnetism of oxides, it is shown how the molecular-field hypothesis can amount for the magnetic properties of ferromagnetics and antiferromagnetics, and for those of non-compensated antiferromagnetics, with which latter materials we are concerned here. A brief description of the spinel lattice is given, and also an account of the crystal chemistry of the spinels, which is necessary to understand the experimental saturation magnetizations discussed. A short survey of methods of preparation is given. The second part discusses the anisotropies, and some of the magnetization processes which influence permeability, and the factors which influence high-frequency permeability and losses. Among these are the ferromagnetic resonance phenomenon and the dimensional resonance and relaxation phenomena. The way in which these factors are influenced by chemical composition and preparation technique is indicated. Finally, a short history of the development of ferrites is given.

2340: C. O. Jonkers: Die Philips Gaskältemaschine (Allgemeine Wärmetechnik **6**, 203-206, 1955).

Abbreviated version (in German) of article published in Philips tech. Rev. **16**, 69-78, and 105-115, 1954/55.

2341: J. S. van Wieringen: Paramagnetic resonance of divalent manganese incorporated in various lattices (Disc. Faraday Soc. No. 19, 1955).

Paramagnetic resonance was observed in powder samples containing Mn diluted in various diamagnetic compounds. The measurements were made at 3.2 and 1.25 cm, at room temperature. Three effects were found: narrowing of hyperfine splitting by covalent bonding and by exchange, and a g-factor somewhat larger than the free-electron value in ZnSe-MnSe and CdTe-MnTe mixtures. The first effect suggests 10-20%, 30%, 35% and 40% covalent bonding in Mn-oxygen compounds, MnS, MnSe and MnTe respectively.

2342: A. A. Kruithof and J. L. Ouweltjes: Colour rendering by de luxe fluorescent lamps (Proc. Int. Comm. Ill. Zürich, 1955).

Description of a method of improving the colour rendering of fluorescent lamps by the addition of a special phosphor. The principles governing the colour rendering are discussed and the application of these principles demonstrated with regard to some newly developed "de luxe" lamps.

2343: J. J. Balder and G. J. Fortuin: The influence of time of observation on the visibility of stationary objects (Proc. Int. Comm. Illn. Zürich, 1955).

Interim report of new measurements, with new apparatus, of the threshold values of visual acuity. Apart from the variables concerned in an earlier investigation by Fortuin (see these abstracts **R 170** and **R 174**), the effect of object observation time was also examined.

2344: H. A. Klasens, P. Zalm and G. Diemer: Characteristics of electroluminescent cells (Proc. Int. Comm. Illn. Zürich, 1955).

A consideration of the chemical nature of an electroluminescent phosphor is followed by a discussion of the electrical and optical characteristics of an electro-luminescent cell, making use of a simple equivalent circuit. On the basis of these considerations, and direct observations, the properties of the individual phosphor crystals are derived. A relation is found expressing the brightness B in terms of the frequency ω and the applied voltage V. A theory is proposed for the mechanism of electroluminescence which explains this relation. Finally, some comments are made on the practical application of electroluminescent lamps.

2345: H. Zijl: Computed coefficients of utilization (Proc. Int. Comm. Illn. Zürich, 1955).

A method is described by which utilization tables can be computed which take account of the nature of the illumination of the working surface and the light distribution over all the surfaces enclosing the illuminated space. It is possible to present the data in the form of simple graphs and tables.

2346: J. B. de Boer and J. F. T. van Heemskerck Veeckens: Observations on discomfort glare in street-lighting; influence of the colour of the light (Proc. Int. Comm. Illn. Zurich, 1955).

Experimental investigation into the effect of the colour of the light on glare in street lighting. Of the various lamps used, the observers showed a definite preference for the "warmer"-coloured light sources. The acceptable luminance (to ensure a certain degree of comfort) was found to be approximately proportional to the square root of the road luminance. Differences in the results from those of other investigators can partly be attributed to differences in the measuring techniques.

2347: H. Bremmer: Diffraction problems of microwave optics (I.R.E.Trans. 4, 1955).

Survey of current methods applied in microwave diffraction theory.

2348: A. A. Kruithof: Chromatic adaption with near white backgrounds (Die Farbe **4**, 147-158, 1955).

Repetition and extension of work carried out earlier in collaboration with P. J. Bouma (Philips tech. Rev. **9**, 257-266, 1947/48).

2349: J. L. Meijering, G. W. Rathenau, M. G. van der Steeg and P. B. Braun, A miscibility gap in the face-centred cubic phase of the copper-nickel-chromium system (J. Inst. Metals **84**, 118-120, 1955/56).

The copper-nickel-chromium system has been studied by metallographic and X-ray-diffraction methods, particular attention being paid to the isothermal section at 930 °C. In addition to the body-centred cubic phase rich in chromium, two face-centred cubic phases have been found to exist, although the binary copper-nickel alloys show complete miscibility.

2350: C. Meyer: Gasentladungslampen und ihre physikalischen Probleme (Physikalische Blätter **12**, 19-28 and 62-73, 1956). (Gas discharge lamps and their physical problems; in German)

Following a short historical review, questions relating to the efficiency of luminous sources, fluorescence, colour rendering and the conditions of operation (striking and stabilization) are discussed. The lamps discussed include sodium vapour lamps, high pressure and super-high pressure mercury lamps, high pressure xenon lamps and electronic flash tubes. The article ends with a table showing the spectral distribution of the light from a number of gas-discharge lamps as compared with daylight and incandescent lamplight.

2351: O. W. Memelink: The distribution of impurity in a semi-infinite solidified melt (Proc. Phys. Soc. **B 69**, 119-120, 1956).

The transient distribution of impurity segregating upon solidification in a semi-infinite melt has been calculated. The expression obtained, which is of closed form, contains the following parameters: the constant velocity of the moving solid-liquid interface, the ratio of impurity solubility in the solid to that in the liquid state and the diffusion constant of the impurity in the liquid state.

2352: J. E. Rombouts and J. Links: The chemical nature of the antibacterial substance present in Aucuba japonica Thunbg (Experientia **12**, 78-80, 1956).

An inactive precursor of the antibacterial substance present in the aucuba plant has been isolated and identified from the seeds. The precursor was identified with aucubin, a glycoside of a furan derivative. From the leaves an enzyme could be extracted which produced from the aucubin (presumably by hydrolysis) the antibacterial substance aucubigenin. The activity spectrum of this substance is described. It is asserted that the antibacterial substance present in crude juices from leaves is identical with the hydrolysis product of aucubin.

2353: H. J. G. Meyer: Interaction of excitons with lattice vibrations in polar crystals I (Physica **22**, 109-120, 1956).

A theory describing slow excitons which interact with longitudinal optical lattice vibrations is developed within the framework of the effective mass approximation. In many respects the theory is very similar to — and makes explicit use of — the ordinary theory of polarons. Interesting differences with polaron theory are introduced by the fact that an exciton has an internal degree of freedom and is as a whole electrically neutral. The effect of lattice vibrations on optical exciton transitions is investigated with the aid of the above theory.

2354: G. D. Rieck: Fragmentation in tungsten crystals (Acta Met. **4**, 47-51, 1956).

Back-reflexion-Laue diagrams were made of W crystals in recrystallized lamp filament wires, using a specially collimated X-ray micro-beam. Differences in orientation of 1-2′ could be observed. Sometimes two crystals with orientation difference of 1°-2° are irradiated at the same time. The crystals often showed Laue diagrams with spots split in the direction of the wire axis, with angular differences of 2′-30′. After cold bending, fragmentation was found, which was more clearly seen after re-straightening, for then the deformation asterism diminished. The fragmentation occurred mainly in two directions. One corresponds with the normal breaking-up of the lattice in the direction of bending. The other was approximately perpendicular to the latter and gave small lines parallel to the wire axis, arising from particles with orientation differences of 1′-10′ and dimensions of about 10 μ. Bearing in mind the split spots occurring in unbent wires and the fact that during drawing contaminations in the wire are stretched out along its length, we may assume a predisposition of the crystals to break along planes parallel to the axis.

2355: W. K. Westmijze: The fundamentals of magnetic recording (T. Ned. Radiogenootschap **21**, 1-15, 1956).

A short description is given of the magnetic recording method in general, and some details are treated more elaborately. For the understanding of the h.f. biasing method use is made of a simplified hysteresis curve. In this way an explanation can be given of some of the peculiarities met in recording and erasing. The recorded signal is attenuated by the demagnetizing field, and during reproduction not all the flux in the tape is reproduced. This is discussed for short, long and intermediate wavelengths. Finally, the factors are surveyed which influence distortion, frequency response, noise and print effect.

2356: G. P. Bakos: Mechanical aspects of magnetic-recorder design (T. Ned. Radiogenootschap **21**, 17-37, 1956).

The quality of magnetic recording is determined to a great extent by the electro-mechanical design of the tape-drive mechanism. The task of this mechanism is to make the tape run with a constant speed and in intimate contact with the magnetic heads. It depends on the constructional details how far this task is fulfilled. This article deals with the various factors which have to be taken into account when designing tape drive mechanisms.